

Reining in assisted reproduction

Fears about the safety of reproductive technologies should be kept in perspective, but more research is needed to assess the risks. In the meantime, its practitioners should learn the virtue of caution.

Fertility specialists derive enormous satisfaction from their work. Whereas many doctors are confronted daily with patients who are struggling against death, the fertility expert's stock-in-trade is the creation of life. It must be uplifting to arrive at work each morning and glance up at a wall filled with smiling family portraits that are the direct result of your professional endeavours.

But this emotional reward, along with the commercial pressures that bear on private reproductive clinics, means that caution is not always the watchword. Many clinics tout for business by quoting their success rate, so they are eager to adopt techniques that boost the chance of a successful pregnancy. At the same time, a desire to help more couples experience the joys of parenthood has led reproductive specialists to adopt more aggressive techniques to treat infertility. Intracytoplasmic sperm injection (ICSI), for instance, in which a sperm cell is injected directly into the egg, is now a routine procedure.

Some researchers are now starting to question the safety of ICSI and other techniques, claiming that they are linked with increased rates of birth defects and rare 'genetic imprinting' disorders (see page 656). This sounds alarming, but it's important to keep the fears in perspective. So far, the few studies done are mostly based on small sample sizes, and in some cases the findings are contested. Other studies on children conceived by assisted reproduction have found no evidence of any problems, and some developmental biologists see no reason to suppose that such techniques should pose a significant risk.

But there are good reasons to urge caution. We know from studies of livestock, where the manipulation of eggs and embryos is often more severe than in human fertility clinics, that such interventions

may be associated with a syndrome in which fetuses grow too large, and may die at around the time of birth. Disturbingly, there seem to be parallels between this condition and the imprinting disorders now being linked tentatively to assisted human conception.

The tendency of human fertility specialists to push the boundaries ever outwards is a further cause for concern. Consider, for example, the sorry tale of cytoplasmic transfer, pioneered by Jacques Cohen of the Institute for Reproductive Medicine and Science of Saint Barnabas in Livingston, New Jersey. By injecting their eggs with cytoplasm from eggs donated by younger women, Cohen enabled some infertile women to have a child. But the technique also seems to heighten the risk of a chromosomal abnormality.

Few would argue that such adventures are desirable, but how can they be curbed? Requiring all new assisted-reproduction techniques to undergo full clinical trials would probably bring progress to an end. But *Nature* has argued previously that there is much to commend the British model, under which a statutory authority regulates fertility clinics and researchers (see *Nature* **420**, 1; 2002). In many countries, clinics in the private sector are given too free a rein.

There is also a need for more research to assess the risks posed by assisted reproduction. This will require further epidemiological follow-up and studies to determine whether the embryos created bear any cellular or molecular abnormalities. The current US administration, unfortunately, is unlikely to provide federal dollars for research on human embryos. For the sake of future generations of assisted-reproduction children, funding bodies in more permissive countries should rise to this important challenge. ■

Anything you can do...

Spy scandals notwithstanding, the party goes on at the two US nuclear-weapons design labs — and so does the backbiting.

The Lawrence Livermore National Laboratory grew out of a lab established in 1952 by a group of dedicated individuals who believed they could do better than Los Alamos, and build a hydrogen bomb. Some of them also believed that Los Alamos was run by communist sympathizers, and testified in Congress to that effect.

The two establishments have had a less-than-cordial relationship ever since. Their rivalry retains a sharp edge: only last year, Livermore staff sabotaged an effort to foist on them a director who had spent most of his career at Los Alamos (see *Nature* **417**, 577; 2002). The rivalry has suited successive US governments just fine, but the latest arena of competition may be less pleasing to Washington. Almost five years after Los Alamos descended into the Wen Ho Lee spy scandal, Livermore has now embarked on a spectacular riposte (see page 651).

Los Alamos may have been content to oversee the bungled investigation of a middle-aged Taiwanese mechanical engineer, struggling to figure out which thermodynamics meetings he had attended in Asia. Livermore, as always, seems to have been up to something more exotic.

According to court documents, the head of security at Livermore (before he resigned last week) enjoyed an on-off relationship over many years with a Los Angeles socialite and Republican fundraiser,

Katrina Leung, who worked for the FBI but is now in prison pending trial on charges that she was also operating as a double agent for Beijing. The FBI's nightmare is that her alleged operation may have enabled China to keep an eye on all of the agency's spying investigations — including those involving leaks from Livermore, Los Alamos and elsewhere — over two decades.

There's no sign that any scientists from Livermore were invited to the splendid garden parties that Leung liked to throw in one of Los Angeles' smartest neighbourhoods. It's not clear that she was particularly interested in the technicalities of nuclear weaponry: she was more of a person person, by all accounts.

Nonetheless, the queue of postdocs keen to further their careers at both Livermore and Los Alamos remains impressive. It isn't just the glamour that draws them, it is the money that Congress continues to pump into each establishment. During the Wen Ho Lee scandal, there was talk that this largesse might dry up; instead, it has multiplied. If, as seems increasingly certain, Livermore's most valuable innovations in the miniaturization of thermonuclear warheads were passed straight to Beijing in the 1980s, that's all the more reason, Congress seems to think, to innovate some more. ■





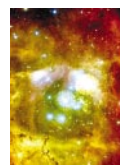
Fighting back
Therapeutic vaccine
shows promise
against HIV
p650



Health alert
SARS brings
Ontario's hospitals
to a virtual standstill
p652



Life on Mars?
NASA gears up for an
in-depth trip to the
red planet
p653



Snap happy
The world's largest
digital camera takes
aim at the sky
p654

Flight records reveal full extent of Agent Orange contamination

Declan Butler

A fresh study of long-forgotten flight records of US military aircraft that sprayed Agent Orange over Vietnam has shed unexpected light on one of the darkest episodes of that conflict.

The study, published on page 681 of this issue, revises previous estimates of the quantity of the herbicide sprayed during the Vietnam War sharply upwards. Together with the earlier use of other herbicides — known as Agent Purple and Agent Pink — it finds that the total amount of dioxin contaminant sprayed in the war was up to four times as great as was previously estimated.

And in a report due to be released on 17 April, the US Institute of Medicine declares that the new work shows that it is feasible to perform large-scale epidemiological studies of the links between the herbicide spraying and the health of the Vietnamese population and US veterans. The institute calls on the US government to support such studies.

The new analysis, performed by a group led by Jeanne Mager Stellman at Columbia University in New York, provides the most detailed and sophisticated computerized maps ever produced of herbicide spraying in Vietnam. For the first time, the authors say, it is possible to calculate an exposure index for individuals and populations that is accurate enough for the epidemiological research that is needed for firm links with health data.

Stellman's team calculated the index by compiling a computer database that overlays staggering amounts of various data, much of it newly unearthed, onto geographical coordinates. The data include flight-path information, the amount and type of agents delivered (including releases caused by leaks, crashes and dumps), troop locations and movements, land features and soil type, and the location of Vietnamese populations.

"I just think, 'Wow', if only this had been around earlier," says Alistair Hay, an environmental toxicologist at the University of Leeds, UK. "It fills in the gaps, giving an accurate and comprehensive picture of where spraying took place."

Like a widely cited 1974 National Academy of Sciences study on Agent Orange, the



Up to 4 million Vietnamese may have been affected by American wartime defoliant spraying.

researchers relied heavily on HERBS, an electronic record compiled by the US military, which includes data on flight paths, herbicide agents and volumes used, and other information on 10,000 spraying missions. But when digging out and cleaning up these data, the team stumbled on valuable archive data that had been overlooked, and discovered a connection that allowed them to extract far more information from existing records.

A major shortcoming of earlier analyses using HERBS was that its spraying data were treated chronologically, leading to mission flight paths criss-crossing the country but without showing patterns of which areas were sprayed over time. The new study's breakthrough came when Stellman, browsing army archives, found another source of information: the daily logs filed by pilots after missions. "It was a sort of 'eureka moment' when I realized that there were 'project' numbers on each of these reports, and that those numbers vaguely resembled a couple of columns that we have never put together before or used," she says. "I just thought, 'Oh my God, look at this!'"

These columns of numbers included a column called 'missionnum', which, when linked with data in the 'military region' column, matched up with project numbers kept in the pilots' logs. The logs contained exactly what HERBS lacked: details of the missions' targets. The link enabled the team

to draw the first maps of spraying patterns in Vietnam that show precise localized target zones, and how much herbicide was sprayed on them and when.

It is possible to tell from the maps, for example, whether individual soldiers or populations were likely to be present in a particular zone on the day of spraying and exposed directly, or whether they arrived later and were exposed indirectly.

The findings are set to provoke a vigorous reaction both from US veterans' groups — some of which are still seeking compensation for exposure to Agent Orange — and the Vietnamese government. Using census data for 20,000 Vietnam hamlets for the first time, the study shows that at least 3,000 of them were sprayed directly, affecting between 2 million and 4 million people. The researchers also plan to publish maps of spraying in relation to US troop positions.

The model used in the study is "a valid means for assessing wartime herbicide exposure in Vietnam veterans", says the Institute of Medicine report, written by a committee chaired by David Hoel, an epidemiologist at the Medical University of South Carolina in Charleston. It commends the team for its "dogged pursuit of historical records" and recommends that the Department of Veterans Affairs, which funded the study, work with other government agencies to support new epidemiological studies of veterans' health. ■



DOD/AP, R.VOGEL/AP (INSET)

Trial suggests vaccines could aid HIV therapy

Erika Check, Washington

A therapeutic vaccine could help patients with HIV to take a break from their exacting drug treatments, according to a prominent immunologist.

Brigitte Autran, a physician at the Pitié-Salpêtrière Hospital in Paris, revealed the results of her recent study on 3 April at a Keystone symposium on HIV vaccines in Banff, Canada.

At the very least, Autran claims, the study offers the best evidence yet for the promise of a therapeutic vaccine in managing HIV. But the small size of the clinical trial has led some researchers to question the findings.

"Our results suggest that a vaccine that does not provide sterilizing immunity could still have an impact on the clinical management of the disease," Autran told the meeting. "This is one of the first proofs of principle supporting the concept of therapeutic immunization."

A therapeutic vaccine would not stop people from contracting HIV. But, in theory, it could be used to boost the immune systems of individuals with HIV who are undergoing 'highly active anti-retroviral therapy' (HAART).

Many HIV patients in wealthy countries are now being prescribed HAART in an effort to delay the onset of AIDS. It suppresses HIV but also causes harsh side-effects, and patients eventually develop resistance to the treatment. Using a therapeutic vaccine to boost the body's immune system in the fight against the virus might allow patients to take a break from HAART. This would provide both a respite from the side-effects and an extension to the useful lifetime of the therapy.

Autran says that her trial is the first to show that a vaccine can do just that. She and her colleagues injected 48 HIV-infected people on HAART with a therapeutic vaccine, which contained HIV genes stitched into a harmless bird virus.

Autran says that more than half of the patients developed immune responses to the vaccine. All of the patients then went off HAART, and doctors measured how long their immune systems could control HIV. Those who developed immune responses to the vaccine managed to stay off HAART for a median of 10 weeks, around 18 days longer than those who did not respond, Autran says. Although she admits that the effect was relatively small, Autran says that it could be amplified with better vaccines.

Other scientists point out that Autran's trial did not include a control group of patients on HAART who did not receive the vaccine, so it is hard to know whether the vaccine actually benefited the patients. Researchers such as Martin Markowitz and

his colleagues at the Aaron Diamond AIDS Research Center in New York have also found that patients on HAART can develop strong immune responses to vaccines. But in the long term, this did not allow the patients to stay off HAART any longer than patients who were not vaccinated (M. Markowitz *et al. J. Infect. Dis.* **186**, 634–643; 2002).

"Autran's data confirm the fact that the immune system can control the virus to some degree," Markowitz says, "but the challenge is whether we can harness the immune system to control the virus over the long term."

This does not invalidate the therapeutic-vaccine approach, Markowitz adds. Like Autran, he believes that better vaccine candidates could translate into stronger results. Both he and Autran used the best vaccines available for their trials, but newer vaccine candidates might prove to be more effective.

Convincing regulators to approve a therapeutic HIV vaccine may be a challenge in itself, however, as it is unclear what data would be needed to indicate efficacy. Most vaccines work by blocking infection entirely, says Alan Landay, an immunologist at Rush Medical College in Chicago. Landay is



Brigitte Autran thinks the vaccines can augment drug therapies for managing HIV.

organizing a meeting in Washington later this month to bring government officials, industry representatives and scientists together to talk about how best to proceed with therapeutic vaccines. ■

Chemists' salaries turn south

Hannah Hoag, Washington

The wallets of newly qualified US chemists are getting lighter for the first time since the mid-1990s. The American Chemical Society's annual salary survey shows that the median salary for PhDs in 2002 has dropped \$2,000, to \$67,500.

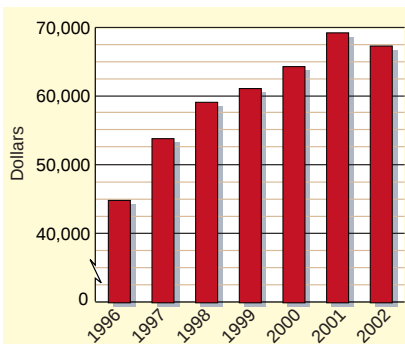
Salaries for newly awarded PhDs have climbed steeply in recent years — they now earn around 50% more than those who qualified six years ago. But this year's fall in starting salaries suggests that the trend is changing, probably because the economic

expansion of the 1990s has come to an end.

Median starting salaries of \$71,000 for PhDs in industry is one reason that, overall, twice as many chemists are employed in the industrial sector as in research institutions. But large chemical companies such as DuPont and Dow aren't wooing newly qualified chemists any more, says Charles Casey, president-elect of the American Chemical Society. New chemists are more likely to start their careers with smaller companies, he says, a move that could account for the dip in salaries.

Hannah Bernstein, an assistant director of the Careers Office at the Massachusetts Institute of Technology (MIT), says that PhDs from MIT who might have gone on to industry or consulting in the past are now pursuing postdoctoral training. According to the society's survey, 47% of new masters were opting for further study in 2002, compared with 27% in 2000.

Bernstein contends that salaries are lower across all scientific disciplines. During the economic boom, companies had to offer large salaries and signing bonuses to be competitive and attract good candidates, she says, but now they offer less. "The overall economy may be the driver," agrees Casey. ■



Average starting salaries for chemistry PhDs in the United States take a dip.

Livermore official linked with alleged spy

Geoff Brumfiel, Washington

The head of counterintelligence at Lawrence Livermore National Laboratory in California has resigned following revelations that he had a sporadic, 15-year affair with an alleged Chinese spy.

William Cleveland, a former special agent with the Federal Bureau of Investigation (FBI), resigned from his position at the nuclear-weapons laboratory on 10 April, after ten years in charge of its counterintelligence office. His resignation preceded press revelations that an official matching Cleveland's description had a relationship with Katrina Leung, a prominent Chinese-American community leader and Republican Party fundraiser. Leung worked for the FBI but is now accused of working as a double agent for China. She was arrested on 9 April and is in prison awaiting trial. Cleveland has not been charged with any offence and is cooperating with the authorities.

"The allegations in the affidavit are personal in nature, and there is absolutely nothing in it to indicate that security was compromised at Livermore," Susan Houghton, a



Katrina Leung:
accused of spying
for China.

spokeswoman for the laboratory, told *Nature*. Without naming Cleveland, Houghton added that the lab has barred the official in question from the campus and impounded his personal equipment pending an investigation. Congressional leaders were briefed on the case last week, but the committees responsible for overseeing the lab have yet to decide whether they will conduct their own investigation.

The main allegations against Leung involve her relationship with FBI agent James Smith, who handled her work for the bureau. It is not clear how much, if any, information allegedly passed by Leung to China involved the Livermore laboratory. But she seems to have some knowledge of the investigation of a scientist there, Peter Lee, who confessed in 1997 to leaking nuclear-weapon and subma-

rine secrets to China. According to the FBI affidavit, a search of Leung's residence in December 2002 turned up a list of telephone numbers for 'Royal Tourist', the codename for the FBI's investigation into Lee.

Cleveland was said to be well respected at the lab, where he trained researchers and managers on how foreign agents might try to befriend them to gain classified secrets. He was also responsible for debriefing scientists who had travelled to sensitive countries or recently hosted foreign visitors. "He was a superstar of counterintelligence at the Department of Energy," says Peter Stockton, who served as a security consultant to former energy secretary Bill Richardson, and now works with the Washington-based Project on Government Oversight.

The latest espionage case could carry adverse consequences for the laboratory or its contractor, the University of California, which is already under fire for its management of Los Alamos National Laboratory in New Mexico (see *Nature* **421**, 99; 2003), and for an earlier, inconclusive spying probe involving Wen Ho Lee, a researcher at Los Alamos. ■

Partners dig deep for ocean-drilling project

Geoff Brumfiel, Washington

Japan and the United States are set to forge ahead with a major drilling collaboration to explore what lies beneath the ocean floor.

The two nations are expected to sign a deal next week that will commit US\$1.2 billion over ten years to the operating costs of the Ocean Drilling Program. And although Europe won't take part in the Tokyo signing ceremony, ocean scientists are optimistic that it will come on board later.

The deal, which succeeds an existing agreement that expires this September, will deploy two ships equipped with advanced drilling technology to take core samples from the ocean floor, including from areas that have previously been inaccessible.

The US National Science Foundation (NSF) and the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT) have each agreed to bear 44% of the estimated \$150-million annual running costs of the programme, and the hope is that Europe will make up the deficit. The agreement marks "a visionary new phase in scientific ocean drilling," says Jeff Fox of Texas A&M University in College Station, the director of scientific operations for the existing international programme.

Japan and the United States are already planning to build the two vessels that will do most of the programme's work. Japan is

spending \$500 million on *Chikyu*, a vessel that will be able to bore up to six kilometres into Earth's crust.

"This new ship will try to understand the physical and chemical environment off Japan's coasts, where many earthquakes occur," says Ted Moore, a marine geologist at the University of Michigan in Ann Arbor, who co-chairs the new programme's scientific review committee.

In addition, Moore says that the ship will allow scientists to learn more about the formation of Earth's crust and to sample the huge fans of sediment that lie in the deep ocean. These sediment fans, which are deposited on the edge of continental shelves, may be one of the most complete sources of information on the planet's climate history.

The United States is planning to spend \$90 million on a more modest drilling vessel, probably a refurbished version of its current ship, the *JOIDES Resolution*. This could be ready next year, whereas the *Chikyu* is expected to enter service in 2006.

A European consortium of scientists sought funds for a full one-third participation in the new programme, but was unable to secure funding under the European Union's Sixth Framework Programme. European researchers are still planning to participate through smaller, targeted drilling projects, says Chris Franklin, head of Earth

science and technology at Britain's Natural Environment Research Council, and chairman of the consortium. "Europe is striving hard to provide a way of funding this science in 2004," he says. ■

► www.oceandrilling.org



Cutting edge: the drill ship *JOIDES Resolution* prepares to take samples from the ocean floor.

Taiwan left isolated in fight against SARS

David Cyranoski, Tokyo

Researchers in Taiwan say they are being shut out of the global investigation into the pneumonia-like disease that is sweeping the world because their country isn't recognized by the World Health Organization (WHO), which is coordinating the study.

Mei-Shang Ho, an epidemiologist at Taiwan's Academia Sinica, is one of several researchers who say they have been denied access to samples and information. Ho wanted more data on severe acute respiratory syndrome (SARS) after it affected two Taiwanese patients in early March. But WHO officials told Ho and other investigators that they should instead approach the People's Republic of China in Beijing, which the organization does recognize.

"Knowing the information is there but not being able to get it is truly frustrating," says Ho, who decided to travel to Beijing last week to get first-hand data on the disease's spread. "We can't get any information from the WHO," says Yuan-Tsong Chen, director of the Institute of Biomedical Sciences at the Academia Sinica in Taipei.

China has never recognized the government of Taiwan, where Chinese nationalists retreated when the communists won control of the mainland in 1949. It considers Taiwan to be its province, and since 1972, when the United Nations admitted the communist government in Beijing, the WHO and other bodies have denied Taiwan membership.

Unable to participate in meetings between the WHO's collaborating centres, Taiwan's researchers must rely on the WHO's website.



"By the time the information is in the public domain, it's probably out of date," says Chen. WHO officials are keen to visit mainland China, but have not asked to visit Taiwan.

Taiwanese researchers say they have also struggled to obtain materials such as diagnostic reagents. To test Taiwan's 23 probable SARS cases for the coronavirus thought to be responsible, researchers wanted to use antibody tests distributed by the WHO. "We were told to go to Beijing," says Chien-Jen Chen, an epidemiologist at National Taiwan University and chair of the island's SARS advisory committee.

The WHO says its hands are tied because Taiwan is not a member. Iain Simpson, a WHO spokesman in Geneva, says that the organization arranged for Taiwan to receive a visitor from the Centers for Disease Con-

trol and Prevention in Atlanta, Georgia, one of its collaborating centres. "There's little more that we can do," he adds.

Government officials are highlighting the situation to promote Taiwan's case for full representation at the United Nations and the WHO. "We share our public-health officials' anger that for political reasons we are denied access to the world health system," said Parris Chang, a legislator with Taiwan's ruling Democratic Progressive Party, on a visit to Washington DC last week. He added that Taiwan is demanding observer status at the World Health Assembly, the WHO's governing body.

Congressman Sherrod Brown (Democrat, Ohio) backed Chang and called for the United States to support the recognition of Taiwan more vigorously. Taiwan's current plight is an "unjustified violation of the basic human right to healthcare", says Brown.

Despite its isolation, Taiwan seems to be dealing successfully with the epidemic. Since mid-March, when the first cases were recognized, it has reported one new case a day, compared with five each day in Singapore and 40 in Hong Kong. There have been no deaths.

Taiwanese researchers have also enjoyed some collaborative success. A Taiwanese group was involved in a study that isolated the coronavirus from patients in Taiwan and elsewhere (T. G. Ksiazek *et al.* *N. Engl. J. Med.* doi:10.1056/NEJMoa030781; 2003).

● Researchers at Canada's Michael Smith Genome Sciences Centre in Vancouver, British Columbia, said on 12 April that they have sequenced the coronavirus that is thought to cause SARS (www.bcgsc.bc.ca). ■

Infection risk puts the brakes on Canada's biomedical research

The epidemic of severe acute respiratory syndrome (SARS) is freezing up biomedical research in Toronto, Canada, as medical-school administrators are forced to reduce access to hospital buildings. And the situation is being eyed nervously by colleagues in the United States and elsewhere, who fear similar consequences as the spread of the mystery illness continues.

"When they started taking our temperatures we knew it was getting serious," says Stephen Scherer, a geneticist at Toronto's Hospital for Sick Children. But for Scherer and other scientists at Toronto's research hospitals, a nurse's check-up as they arrive each day is just the start of the problem.

To contain the spread of the disease from sufferers admitted to the hospital and staff working there, all non-essential staff at Scherer's institute and the nearby Mount Sinai Hospital are being encouraged to stay at home.

Researchers determined to get to work must join queues at the hospital's entrances, pass a SARS screening test and wear a surgical mask

while in the hospital. Meetings of researchers are banned, even outside hospital grounds.

"There's been massive disruption," says Michael Tyers, a cell biologist at Mount Sinai's Samuel Lunenfeld Research Institute. "Many of the projects in my lab have been compromised."

Until researchers figure out the exact cause of SARS, and how it spreads, disruption is likely to continue. The needs of researchers will take second place to the need to contain the disease, says infection-control specialist Mark Loeb of McMaster University in Hamilton, Ontario, who is advising Toronto's hospitals on controlling SARS.

Many researchers who have no direct contact with patients still like to rub shoulders with clinical staff, and administrators think that this helps to assure the research's clinical relevance. But the SARS outbreak is giving some researchers second thoughts. "I'm starting to question the point of working in a research hospital," says Tyers.

Cell-signalling researcher Tony Pawson, research director of the Samuel Lunenfeld



Warning signs: access to the Hospital for Sick Children in Toronto is temporarily restricted.

institute, says he is optimistic that the SARS outbreak will abate before researchers walk out. "But it raises a lot of issues about hospitals being used as thoroughfares," he says. **Tom Clarke**

NASA homes in on sites for Mars exploration

Tony Reichhardt, Washington

After more than two years of deliberation, a committee of planetary scientists has picked the landing sites for NASA's two Mars Exploration Rovers, due to launch between May and July this year. The two locations — Gusev Crater and Meridiani Planum — have been chosen because of the high likelihood that water, and perhaps life, once existed there.

The rovers will use parachutes and protective airbags to bounce down on to the martian surface next January. Once deployed, they will spend at least three months rolling slowly from rock to rock, taking panoramic photographs and examining the local geology. Each rover is equipped with a high-resolution camera, a diamond-tipped grinder to expose fresh surfaces on dusty rocks, a microscopic imager to examine such surfaces, and three spectrometers to analyse the rocks' chemical make-up and mineral content.

Gusev's appeal has been obvious since the Viking orbiters took the first close-up pictures of Mars in the 1970s. The ancient impact crater shows strong evidence of once having been flooded with water because its bottom is flat and relatively smooth, like that of a lake bed, and a dry river channel seems to run into the structure at one end. Scientists first scrutinized the Meridiani site in 2000, when NASA's

orbiting Mars Global Surveyor detected large deposits of the iron oxide mineral haematite. As most of the processes that form this mineral on Earth require water, its presence hints that water may have been present at the site in the past. Water is considered essential for life, and finding evidence of past biology is a priority for Mars exploration.

When the landing-site selection committee began work in 2000, it had a list of 155 possibilities. These were quickly whittled down to seven top contenders, says John Grant, a planetary scientist at the Smithsonian Institution's National Air and Space Museum in Washington DC, who co-chaired the panel. Gusev and Meridiani have long been front-runners in terms of scientific merit, but they also had to satisfy several engineering constraints. The elevation of the landing site cannot be too high, for example, as the atmosphere needs to be dense enough for the parachutes to work. Boulders are a landing hazard and smaller rocks can impede the rovers, so the site must be reasonably free from such obstacles.

Questions still remain over another constraint. Project scientists will continue to refine their predictions of windiness at the Gusev site before NASA commits to landing there. NASA's last Mars lander — the 1997 Pathfinder mission — landed early in the



The wild rover: NASA is set to launch two probes to scour the surface of Mars for signs of life.

JPL/NASA

martian morning, when it was least windy. But to make sure that signals are sent to Earth during landing, the new rovers will touch down in mid-afternoon, when the winds are typically stronger and could tear the airbags designed to cushion their landing, says Grant.

NASA can make final decisions on the landing sites as much as a month after the first launch, currently scheduled for 30 May. The second rover will launch in late June. Both rovers will be given names before launch, based on suggestions from schoolchildren. ■

Polish science academy prepares for radical overhaul

Quirin Schiermeier, Munich

Winds of change are set to blow through the Polish Academy of Sciences, according to the plant biologist freshly elected to run the organization.

The academy — like many of its eastern European counterparts — oversees a large but somewhat dilapidated national network of research institutes in a range of disciplines. And although Poland has the largest and one of the strongest economies in the region, critics say that parts of the academy have not changed much since communist rule ended in 1989.

But this month Andrzej Legocki, a plant biologist and former director of the academy's Institute of Bioorganic Chemistry in Poznań, takes over as president. He says that he will commission an external quality review of all 80 of the institutes — and predicts that around ten of them could close once the review is completed later this year.

Legocki promises to tackle the academy's main problems head-on, including low output, ageing research staff and lack of contact with university research.

"I am determined to drastically change the situation," he says. "First we shall



A fresh start: Andrzej Legocki aims to foster competition at the Polish Academy of Sciences.



recognize what is worth maintaining, then we will merge groups and institutes, and put them under new, younger leadership." At 63, Legocki is himself the youngest president in the academy's 50-year history.

Poland is the largest of ten countries in central and eastern Europe that will join the European Union next year, but it has been slower than some others, notably Hungary and the Czech Republic, to reform its scientific infrastructure (see *Nature* 421, 471–472; 2003).

Jerzy Langer, a solid-state physicist at the academy's Institute of Physics in Warsaw and vice-president of Euroscience, a

grassroots organization of European scientists, says that Polish science is in need of a drastic overhaul.

"Poland is simply too poor to subsidize an oversized, mediocre science base," he says.

The academy employs about 4,000 scientists, mostly in permanent positions. Critics say that their lifelong job security gives them little incentive to be competitive. And Legocki concedes that the academy lags behind in many fields.

To boost the position of young scientists in the academy, Legocki pledges that all institutes that do not already have a graduate student programme will establish one, along with fixed-term postdoctoral positions for young researchers. "Young scientists at our institutes will be under my personal patronage," he promises.

"Legocki is the right man at the right time," says Jacek Kuźnicki, a neuroscientist and director of the International Institute of Molecular and Cell Biology in Warsaw. "But for any reform to be successful it is vital the older generation accepts that competition is an essential part of the game." ■

► www.pan.pl/english/index1.htm

Baghdad looters rob museum of priceless ancient treasures

Baghdad Last week's ransacking and plundering of Iraq's National Archaeological Museum, one of the most important museums of Middle Eastern antiquities, may have robbed the world of some of its most important treasures. These include one of the oldest known clay tablets bearing cuneiform script — probably the first written language — which had not yet been deciphered.

Iraq was the home of the Sumerian civilization — one of the oldest known to archaeologists — and has been the subject of extensive research over the past century. The national museum and its stores were packed with artefacts, many of which had not yet been examined. More than 170,000 items are estimated to have been stolen.

The museum was reopened only three years ago after repairs following damage incurred during the last Gulf War, when some 4,000 artefacts were also stolen. Other stolen items include bronze statues, religious statuettes, Sumerian gold jewellery and a 4,000-year-old harp.

Museum staff believe that, besides the local looters, specialists are targeting and stealing items of particular historical value.

Scientists blast European vote to limit stem-cell work

Brussels A European Parliament vote to restrict research on human embryonic stem cells may go beyond the parliament's remit, according to sources in the European Commission.

The proposed new rules, which would seem to prohibit therapeutic cloning, were inserted by a German-led cross-party group of politicians as an amendment to a directive on transplantation. The directive was approved by the parliament on 10 April, but health commissioner David Byrne, who proposed the plan with the aim of controlling the quality and safety of transplant tissues, said last year that the intention was not to comment on the type of cells involved. Sources at the commission confirm that ethical issues of this type cannot be addressed by the European Parliament, and have to be resolved by national governments.

The vote has precipitated a storm of protest from European scientists. "A small minority want to overturn the rights of individual member states to make democratic decisions about human embryonic stem cells," says Robert May, president of Britain's Royal Society. The amended directive will next be considered in June, at the Conference of European Health Ministers in Oslo, Norway.

Stuffed Dolly not starved of attention

Edinburgh Even in death Dolly continues to hog the limelight. The famous sheep — cloned from an adult cell by a research team led by Ian Wilmut (pictured) — went on display last week at the Royal Museum in Edinburgh.

Dolly's skin has been pickled, tanned, washed and stretched over a fibreglass frame. Soft tissue and muscles have been replaced with plasticine. Her nose has been touched up with pink paint, and her eyes replaced with glass replicas.

"We had to try about four or five pairs before we got the right shade of yellow," says Andrew Kitchener, curator of birds and mammals at the National Museums of Scotland. She was also "extremely bouffant-looking after the process, so we had to matt her fleece up a bit", he adds.



Britain shuts door on extra neutron source

London News of the almost certain demise of a planned European neutron source was tempered last week by funding boosts to two existing facilities.

The UK Strategy for Neutrons, released by the Council for the Central Laboratory of the Research Councils on 10 April, confirmed that Britain will not support the proposed European Spallation Source. The decision, which has been anticipated for several months (see *Nature* **421**, 563; 2003), may be the end for the facility, which was designed to be the most powerful neutron source in the world.

In the short term, Britain will instead increase its commitment to the Institut Laue-Langevin's neutron facility in Grenoble, France. It will also invest in its own neutron source at the Rutherford Appleton Laboratory near Oxford, where more than £100 million (US\$157 million) will be spent over the next five years to build a second neutron target. Researchers use neutron sources to study the structure of matter, and the new target will allow them to focus on areas such as colloids and biomolecules.

Human genome finished in time for DNA's golden jubilee

Washington The curtain finally fell on the Human Genome Project on 14 April, with the announcement that the sequencing of the three billion bases of human DNA is complete. The 15-year project was finished just in time for the celebrations of the 50th anniversary of the discovery of DNA's structure.

The sequence is finished according to the standards agreed upon by the International Human Genome Sequencing Consortium — there is no more than 1 error per 10,000 bases, and no gaps except for those that cannot be closed with existing technology. "Finishing off those recalcitrant gaps will be done in an individualized tinkering fashion,"

says Francis Collins, head of the project and director of the National Human Genome Research Institute in Bethesda, Maryland.

The project published its draft sequence in 2001, at the same time as one produced by the private company Celera. Only the public project went on to finish the sequence. Further analyses of the finished sequence are expected to be published over the course of the year.

Smile, it's the world's biggest digital camera

Hawaii The world's largest digital camera was unveiled last week at the Canada-France-Hawaii Telescope atop Mauna Kea in Hawaii.

The 340-million-pixel, US\$100-million MegaPrime camera is roughly 100 times more powerful than a commercial digital camera, say its creators. The telescope's unusually broad field of view will allow the device to photograph huge panoramas.

"This camera will allow us to do some unique stuff," says Christian Veillet, executive director for the telescope. For example, researchers plan to use the camera to look for the distant explosions of dying stars and also to observe the Kuiper belt, a ring of asteroid-like objects beyond the orbit of Neptune.

♦ www.cfht.hawaii.edu



One of the MegaPrime camera's first starry shots.

M. McDONALD/PA

CANADA-FRANCE-HAWAII TELESCOPE

Seeds of doubt

One-horse race: in many current fertility treatments, a single sperm is injected directly into the egg. The success rates are impressive, but some experts are concerned about the technique's safety.

Questions are now being asked about the safety of some of the techniques used to overcome human infertility. Kendall Powell examines whether the health of test-tube babies is at risk.

Kelly and Mike were desperate to have a child, and endured nine rounds of traditional *in vitro* fertilization (IVF) treatment at clinics in their native Australia. Over the course of two years, each attempt ended in failure and heartbreak. Then they were given the chance to participate in research to test a new reproductive technology in the United States. The method, which involves growing IVF embryos in culture for five days to increase the chance that they will implant in the womb, had proved safe and effective in mice and cows. But, as Kelly and Mike were informed, its safety in humans couldn't be guaranteed.

It was a chance they were willing to take. Nine months after their trip to the Colorado Center for Reproductive Medicine in Denver, Kelly gave birth to twins — a girl and boy — who both seem perfectly healthy. Their smiling family portrait is one of scores decorating the walls of the Denver clinic.

Although Kelly and Mike were fortunate to attend an academic centre that makes research into the safety of IVF a top priority, other couples are served by commercial clinics where the bottom line — abetted by a genuine desire on the doctors' part to help patients experience the joys of parenthood — may mean that the potential risks aren't so carefully explained. "There's a commercial aspect on one hand and couples' desperation on the other," says Robert Winston, a specialist in reproductive medicine at Imperial

College, London, who last year wrote a review of the safety of assisted reproduction¹. "That's quite a risky combination and not a very good way of conducting responsible medicine."

Over the past year or so, evidence has begun to emerge that suggests assisted reproduction is associated with increased risks of birth defects, low birth weight, genetic disorders, and maybe even cancer. Although the findings obtained to date are far from conclusive, there is a growing feeling that the regulation of fertility clinics may need to be tightened up, and a more rigorous system for monitoring the health of IVF babies put in place. "There is such a dearth of information," says Laura Schieve, an epidemiologist at the US Centers for Disease Control and Prevention in Atlanta, Georgia.

A helping hand

Since the first test-tube baby, Louise Brown, was born in Britain in 1978, more than a million children have been brought into the world through assisted reproductive technology. The original IVF technique, which involved mixing eggs and sperm in the lab and then implanting the resulting embryos into the womb, was developed to help women with blocked fallopian tubes. Early follow-up studies of IVF children, conducted in several countries, suggested that the technique was safe²⁻⁵.

Emboldened by this success, reproductive specialists turned to progressively more

interventionist techniques to treat infertility. Women have been prescribed drugs to boost ovulation, and eggs and sperm have been manipulated in various ways to ensure that fertilization occurs. One technique in particular — known as intracytoplasmic sperm injection, or ICSI — was quickly taken up because it allowed men who can't produce healthy, mobile sperm to become fathers. Many reproductive techniques were first perfected in animals, but ICSI does not work in cattle or mice. It was discovered in 1991 in a Belgian IVF lab, when a technician tried to place a sperm cell just under the egg's protective coat and accidentally injected it all the way into the egg⁶. Today, ICSI accounts for nearly half of all assisted reproductive treatments in the United States.

For some observers, the rapid adoption of such an aggressive technique raised nagging concerns. Might the injection damage the egg's machinery for cell division, for instance? And is it wise to bypass natural selection in such a determined way? At the very least, the technique will pass infertility problems down to the next generation. It is also possible that the defects that cause male infertility are just part of a wider spectrum of genetic abnormalities carried by some faulty sperm.

"This was a no-holds-barred, 'you-will-fertilize' kind of thing," says Jennifer Kurinczuk, an epidemiologist at the University of Leicester, UK, one of several researchers who was prompted by the widespread uptake of

ICSI to take another look at the safety of assisted reproduction. Kurinczuk and her colleagues used data from the Western Australia Reproductive Technology Registry to look at the risk of birth defects diagnosed by the age of one. In a study published in March 2002, they found that the incidence increased from about 4% in a control group of naturally conceived infants to about 9% in children conceived either by conventional IVF or by ICSI. Even after adjusting for factors including the mother's age and prior births, the risk of birth defects seemed to be doubled for both assisted-reproduction groups⁷.

In a study published back-to-back in the same journal, Schieve and her colleagues mined US data and found that infants conceived by assisted reproduction were two-and-a-half times as likely as the general population to have a low birth weight — and this wasn't simply because there were more multiple births among the assisted-reproduction group⁸. Given that babies with low birth weights are more likely to die in their first 12 months, are more susceptible to infections, and often have respiratory problems, this was clearly a worrying finding.

Defect dispute

These two papers caused a flutter of alarm, but attracted criticism from practitioners of assisted reproduction. The studies, IVF clinicians said, had design flaws, including their retrospective nature. Gleaning data from patients' records can allow biases to creep into the sampling — a prospective study that follows outcomes for a group of patients is regarded as more objective. Kurinczuk's report, in particular, was questioned because her figure for birth defects among the naturally conceived population was about twice that generally accepted as the norm. David Adamson, director of Fertility Physicians of Northern California, a clinic in Palo Alto, finds the results for assisted reproduction hard to believe. "Almost one in ten babies with a defect?" he asks. "It's just not true."

The disagreement stems in part from differences in the definition of what constitutes a birth defect. Kurinczuk, for instance, uses criteria borrowed from the British Paediatric Association that include problems such as an undescended testicle. Two ongoing prospective studies, one in Belgium⁹, the other across six European countries, are monitoring ICSI children and have found no evidence of any worrying trends. But when Kurinczuk reanalysed the Belgian data using her criteria, she found more than twice as many birth defects as had previously been reported¹⁰.

While the arguments about birth defects rumble on, the debate over the safety of assisted reproduction has now spread to rare disorders affecting genetic imprinting — the process that, early in development, can silence either the copy of a gene inherited from your mother, or that from your father.

We are using humans as guinea pigs.

Kelle Moley

Andrew Feinberg of Johns Hopkins University in Baltimore, Maryland, and Michael DeBaun of Washington University in St Louis, Missouri, study children with Beckwith-Wiedemann syndrome, which occurs in 1 in 15,000 births, and results from faulty imprinting. Children with this syndrome are larger at birth than normal, have enlarged organs, and are at increased risk of developing certain cancers. Feinberg and DeBaun were struck by the fact that 4.6% of their registered patients were conceived by assisted reproduction, a high proportion given that less than 1% of the general population is conceived artificially¹¹. DeBaun believes that the similarities to 'large offspring syndrome', sometimes seen in sheep and cattle produced by assisted reproduction¹², are too obvious to overlook. In fact, one of the genes implicated in the oversized livestock, *IGF2*, is also affected by faulty imprinting in Beckwith-Wiedemann syndrome.

An independent British study has since confirmed the association between Beckwith-Wiedemann syndrome and assisted reproduction¹³, and three cases of an even rarer imprinting disorder, called Angelman syndrome, have been discovered among children conceived by ICSI^{14,15}. Most recently,

five Dutch children conceived by assisted reproduction have been diagnosed with retinoblastoma, a childhood cancer of the retina that occurs in just 1 in 17,000 births¹⁶. Given this incidence, one would expect to find only one child with the condition in the Netherlands who was conceived artificially.

Catalyst for action

These studies into imprinting disorders and one rare form of cancer are very small. But the fact that these diseases have known genetic causes has led even those who dismiss the warnings about birth defects to accept that they warrant further investigation. "We'd be foolish to think there are no problems," says Adamson. "These studies absolutely require that we look into it."

That caution stems in part from a recent experience in which the US Food and Drug Administration (FDA) was forced put the brakes on a new assisted reproductive technique. The procedure, called cytoplasmic transfer, was used to 'rejuvenate' eggs from older mothers. It involved injecting into the eggs a tiny amount of material, including the mitochondria that 'burn' sugars to create cellular energy, from an egg donated by a younger woman. The hope was that the younger woman's more vigorous mitochondria would supplement the rapidly developing embryo's energy needs. Although it seemed to help older women to get pregnant using their own eggs^{17,18}, the procedure also dramatically increased the risk of Turner's syndrome, an X-chromosome defect that can cause miscarriage in early pregnancy.



Robert Winston is uneasy about some of the commercial forces that are driving assisted reproduction.

P. TWEDD/SPFL

That was enough for the FDA to step in, and in July 2001 it ruled that anyone wanting to pursue the technique would have to seek permission to perform a full clinical trial as if they were investigating a new drug.

Some observers believe that this ought to become standard practice for any new technique of assisted reproduction. And against this background of heightened concern, even techniques that were introduced precisely to improve the safety of IVF are coming under closer scrutiny. The embryo-culturing method that allowed Kelly and Mike to achieve their dream of starting a family is a case in point. It was developed in the hope of easing the practice of transferring several IVF embryos into the womb to maximize the chance that one of them will implant. Transferring several embryos greatly increases the likelihood of multiple pregnancies, which can be extremely dangerous for both mother and fetuses. For this reason, Britain's Human Fertilisation and Embryology Authority (HFEA) allows clinics to transfer only two embryos at a time under normal circumstances. But in some countries, such as Russia, Greece and Spain, four or more embryos are still transferred in many cases.

Cultured approach

At the Colorado Center for Reproductive Medicine, David Gardner and Michelle Lane have shown that it is possible to obtain much higher success rates for implantation if embryos are cultured in the lab for five days instead of the conventional two or three, using a succession of culture media to mimic the changing conditions experienced by a newly fertilized embryo as it drifts towards the uterus¹⁹. Encouraged by these results, about one in five US fertility clinics has already adopted the technique, with some using it exclusively. But although Gardner and Lane's goal is to make IVF an efficient technique even when only one embryo is transferred, thus completely eliminating the risk of multiple pregnancies, for the time being the culturing technique is still being used in conjunction with the transfer of two or more embryos.

Some researchers fear that increasing the duration of culture, and exposing embryos to different culture media, might endanger early development. Experiments in mice, for instance, have shown that even subtle changes in the ingredients of culture media can dramatically alter the activity of

David Adamson believes that assisted reproduction is safe, but wants to see more research.

imprinted genes²⁰. "We are using humans as guinea pigs," suggests Kelle Moley, who studies pre-implantation mouse embryos at Washington University in St Louis. Because the formulas for human culture media are proprietary and funding for human research is scarce, she warns, studying the effects of culture conditions on the health of IVF babies is nearly impossible. "There is a small number of companies making culture media and their compositions change sporadically," Moley says.

Gardner and Lane counter that their animal experiments, which encompass more than 6,000 embryos, have yielded no evidence of any problems²¹. They have also looked at the birth weights of babies who were cultured as pre-implantation embryos for five days, rather than three, and found no differences that would suggest the technique is causing a human equivalent to large offspring syndrome²².

Uncertain future

Despite the recent upsurge of concern, IVF practitioners stress that most babies conceived by assisted reproduction seem to be perfectly healthy—and they argue that all the evidence suggests that infertile couples are willing to accept small risks in order to have their own children. But with Louise Brown having yet to reach her 25th birthday, other experts warn against complacency. It may be extremely unlikely, says Alastair Sutcliffe, a paediatrician at London's Royal Free Hospital, but there is still the possibility that IVF could cause subtle genetic or cellular changes that might manifest as cancer or some other disease only later in life. "It

could be a disaster waiting to occur," he says.

Given the existing worries, and the potential for further unpleasant surprises, some researchers are calling for increased funding for projects to investigate the basic biology of fertilization and implantation, investigations of



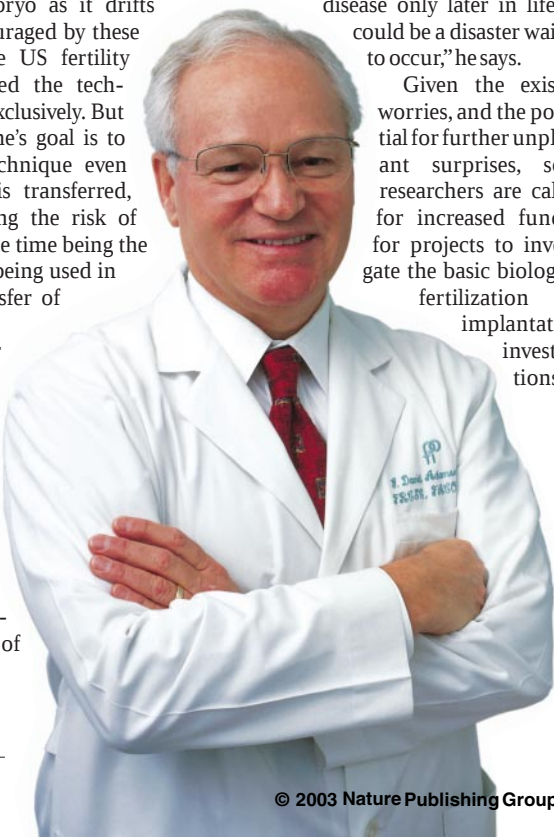
Culturing embryos such as this for five days boosts the chances of a successful pregnancy.

the effects on eggs and embryos of being manipulated in the lab, and bigger and better epidemiological studies. Done properly, the latter will require the prospective follow-up of many thousands of children, collecting standardized data on parental age and medical history, type of infertility, and myriad details about the specific form of assisted reproduction used. Ideally, the children involved would be monitored at regular intervals throughout their lives.

Whether the money to conduct such mammoth studies will be forthcoming remains unclear. But those charged with regulating the field argue that our current state of relative ignorance shouldn't be allowed to persist. "For the benefit of future patients and offspring, it is hugely important to improve the evidence base," says Suzi Leather, who chairs Britain's HFEA.

Kendall Powell is a science writer in Broomfield, Colorado.

- Winston, R. M. L. & Hardy, K. *Nature Cell Biol.* **4**, (Suppl. 1) S14–S18 (2002); *Nature Med.* **8**, (Suppl. 1) S14–S18 (2002).
- Saunders, D. M. & Lancaster, P. *Am. J. Perinatol.* **6**, 252–257 (1989).
- MRC Working Party on Children Conceived by *In Vitro* Fertilisation *Br. Med. J.* **300**, 1229–1233 (1990).
- Friedler, S., Mashiach, S. & Laufer, N. *Hum. Reprod.* **7**, 1159–1163 (1992).
- French *In Vitro* National Fertil. Steril. **64**, 746–756 (1995).
- Palermo, G., Joris, H., Devroey, P. & Van Steirteghem, A. C. *Lancet* **340**, 17–18 (1992).
- Hansen, M., Kurinczuk, J. J., Bower, C. & Webb, S. N. *Engl. J. Med.* **346**, 725–730 (2002).
- Schieve, L. A. et al. *N. Engl. J. Med.* **346**, 731–737 (2002).
- Bonduelle M. et al. *Hum. Reprod.* **11**, 1558–1564 (1996).
- Kurinczuk, J. J. & Bower, C. *Br. Med. J.* **315**, 1260–1265 (1997).
- DeBaun, M. R., Niemitz, E. L., & Feinberg, A. P. *Am. J. Hum. Genet.* **72**, 156–160 (2003).
- Sinclair, K. D., Young, L. E., Wilmut, I. & McEvoy T. G. *Hum. Reprod.* **15**, (Suppl. 5) 68–86 (2000).
- Maher, E. R. et al. *J. Med. Genet.* **40**, 62–64 (2003).
- Cox, G. F. et al. *Am. J. Hum. Genet.* **71**, 162–164 (2002).
- Ørstavik, K. H. et al. *Am. J. Hum. Genet.* **72**, 218–219 (2003).
- Moll, A. C. et al. *Lancet* **361**, 309–310 (2003).
- Cohen, J. et al. *Mol. Hum. Reprod.* **4**, 269–280 (1998).
- Barritt, J. A., Brenner, C. A., Willadsen, S. & Cohen, J. *Hum. Reprod.* **15**, (Suppl. 2) 207–217 (2000).
- Gardner, D. K., Lane, M. & Schoolcraft, W. B. *J. Reprod. Immunol.* **55**, 85–100 (2002).
- Doherty, A. S., Mann M. R., Tremblay, K. D., Bartolomei, M. S. & Schultz, R. M. *Biol. Reprod.* **62**, 1526–1535 (2000).
- Gardner, D., Lane, M., Stevens, J. & Schoolcraft, W. *Fertil. Steril.* **78**, (Suppl. 1) S8 (2002).
- Gardner, D. & Lane, M. *Reprod. Biomed.* (in the press).



Worlds apart

Our knowledge of planets outside our Solar System has been transformed in the past few years. But these new-found worlds don't look much like our planetary neighbours, and no one is quite sure why. Dan Falk investigates.

Less than a decade ago, planetary scientists were working with a tiny data set: the nine members of our Solar System. But the past few years have been a boom time for planet hunters — more than 100 planets orbiting other stars have now been logged. As new detection methods come into use, this tally is certain to climb higher.

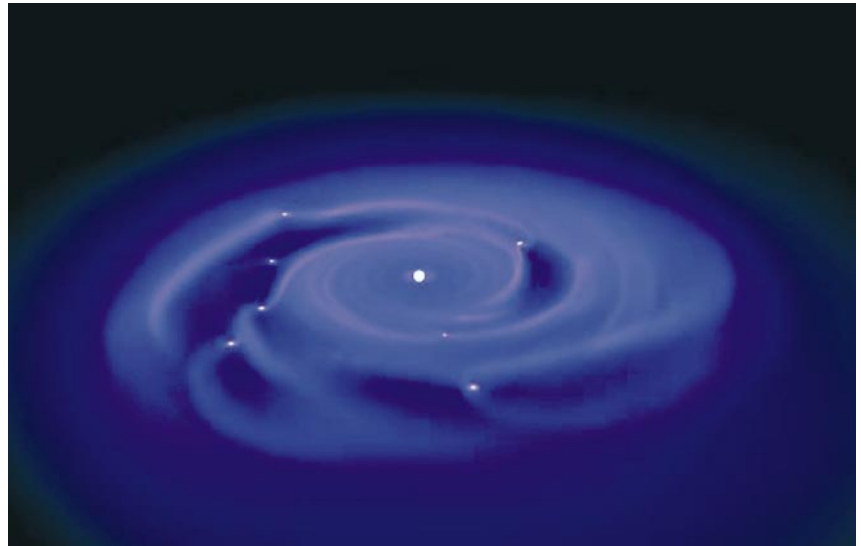
Not everyone is celebrating, however. Extrasolar planets have peculiar properties, and our understanding of how planets form, which was incomplete even before the new data became available, now looks even shakier. The newly discovered bodies have strange, highly elliptical orbits. They are also far closer to their stars than equivalent planets in our Solar System. Amid the thrill of discovery, planetary scientists are wondering how to make sense of the processes that shaped these strange new worlds.

The main method of planet detection is the radial-velocity survey. As a planet orbits a star, its gravity makes the star wobble, causing a periodic change in the spectrum of the star's emitted light from which astronomers infer the planet's presence. This technique will soon be joined by two more: astrometry, in which the star's wobble is measured by tracking its position, and the transit method, in which the passage of the planet in front of the star is observed. Both are still in their infancy — only one planet has been spotted in transit studies, and none has yet been definitively identified by astrometry.

In terms of mass, the new planets are similar to Jupiter, weighing between one-tenth and ten times as much — the majority fall between 0.75 and 3.0 jovian masses. Measuring size is more difficult, as only transit studies can provide information on the object's radius. The planet observed using the transit method — an object orbiting a star in the constellation of Pegasus¹ — is slightly larger than Jupiter.

But that's where the similarities end. The orbits of most extrasolar planets follow elliptical paths, in contrast to the near-circular orbits of our Solar System's giant planets. They also orbit much closer to their parent stars, most at a distance of less than 2 astronomical units (1 AU being the distance between Earth and the Sun), compared with more than 5 AU for Jupiter.

It is these properties that seem to defy pop-



Cloudy picture: computer simulations have yet to nail down the finer points of planetary evolution.

ular models of planetary formation. The two main theories each start with a slowly spinning ball of gas. The hot, central part becomes a star, while the material farther out is flattened by its rotation into a cloud known as a protoplanetary accretion disk. This provides the raw materials from which planets form.

Rocky start

From there on, the process is open to debate, with the answer partly depending on the size of the disk. The core-accretion model, which dates from the 1960s, argues that planets start life as small chunks of rock, dust and sand-grain-sized debris that come together through collisions. As the rocky core grows, its gravitational pull scoops up more dust and gas from the disk. If the core is heavier than a few Earth masses, it accretes enough gas over a few million years to become a gas giant like Jupiter and Saturn. Less-massive cores result in rocky planets like Earth.

This model ran into problems even before extrasolar planets were identified. For one thing, it seems to take too long. Accretion disks are thought to evaporate within a million years or so, probably as a result of the stream of electrically charged particles that all stars emit, or of bombardment from high-energy ultraviolet

photons from other nearby stars.

The main rival theory, which also surfaced in the 1960s, avoids this problem. Known as the disk-instability model, it proposes that, in larger disks, patches of denser gas can form and pull in more gas — leading, in some cases, to a sudden collapse that forms one or more planets. Such collapses do not occur in the core-accretion model, either because the disk is not large enough to produce them, or because any small instability that forms will tend to spread throughout the disk, restoring stability.

Planets are thought to form more rapidly in the disk-instability scenario. Last autumn, Lucio Mayer, a theoretical astronomer then at the University of Washington in Seattle, described high-resolution computer simulations of protoplanetary disks using the disk-instability model². Together with colleagues elsewhere in North America, Mayer showed that giant planets could form in as little as 1,000 years. The difference in planet-forming rates is probably the most important distinguishing characteristic between the two models, and is a boost for the disk-instability idea, says Alan Boss, a theoretical astrophysicist at the Carnegie Institution of Washington.

Others urge caution. Jack Lissauer, a planetary scientist at NASA's Ames Research Center in Moffett Field, California, says that



Starry eyes: the Sub-Millimeter Array in Hawaii will peer into the hearts of distant planetary systems.

the resolution of the computer models is still too poor to give conclusive results. Perhaps more importantly, the new data on extrasolar planets do not sit happily with either theory. The models have trouble explaining, for example, why Jupiter-sized planets are created rather than brown dwarfs — objects that are intermediate in size between planets and stars. “You would expect the mass of planets to range from Jupiter mass up to stellar masses,” says Douglas Lin, an astrophysicist at the University of California, Santa Cruz. There ought to be just as many brown dwarfs as Jupiters orbiting Sun-like stars — something that observations have not turned up.

Inner workings

Other aspects of the new data are causing problems for both models. Neither, for example, accounts for the proximity of the extrasolar planets to their stars. There isn't much material in the inner region of the disk, and the particles there should have enough energy to resist clumping. The solution, astronomers suggest, is that giant planets form farther out and then migrate inwards³ as a result of interactions between the disk and the planet. The mechanism differs in the two models, but the end result is that young planets sail through the disk towards the star.

But this raises another question: what stops the planet from ploughing into its parent star? Several mechanisms have been suggested. One option is that the migration ends when the disk evaporates — but it's not clear whether this can happen quickly enough, as migration occurs on a roughly million-year time scale. Another option is that the planet's gravitational pull distorts the shape of the star, and that this in turn affects the pull of the star on the planet in such a way as to balance the planet's inward movement. Finally, it could be that the star's magnetic field clears out the inner disk by repelling electrically charged particles. In

this situation, says Boss, the inner 0.5 AU of the disk would be empty — and few extrasolar planets have orbital radii much smaller than this. “It's attractively simple,” says Boss.

Such explanations are plausible, but there is no way of knowing which is correct. Even if this issue is resolved, it is still unclear whether planets form by disk instability or by core accretion before they begin their migration. And on top of that, astronomers are struggling to explain why so many extrasolar planets follow elliptical paths, as both formation models predict roughly circular orbits. The best explanation so far proposed is based on the gravitational tug-of-war between different planets in a multi-planet system.

The data needed to eradicate this and other uncertainties could come from a series of new observatories that will allow researchers to study protoplanetary disks, perhaps catching planetary systems in the act of formation. Telescopes that are sensitive to millimetre-range electromagnetic radiation are crucial, as these wavelengths pass through the gas and dust surrounding such



Alan Boss: eagerly awaiting new planetary data.

systems, allowing astronomers to see into the heart of the disk. And although a host star is brighter than its disk across the entire spectrum, the star's radiation is less obscuring at millimetre wavelengths than at optical ones.

Two new telescope arrays operating at these wavelengths will soon start taking readings. The Sub-Millimeter Array on Mauna Kea in Hawaii should get to work later this year, and the Combined Array for Research in Millimeter-wave Astronomy, to be located at a high-altitude site in California, could come online by 2005. But of all the detectors being planned, the most ambitious is the Atacama Large Millimeter Array (ALMA), a joint effort by Europe, Japan, the United States and Canada.

Array of hope

ALMA will use 64 receiving dishes, each 12 metres across, perched high in the Andean plateau in Chile — this array will be equivalent to a single dish 14 kilometres across. Water vapour in Earth's atmosphere is the main enemy of such observations, but the Atacama desert is one of the driest locations on the planet. “It's arguably the best site one can send humans to on a routine basis,” says Lee Mundy, an astronomer at the University of Maryland in College Park, and a member of ALMA's scientific advisory committee.

At one-millimetre wavelengths, ALMA will have a resolution of 0.1 arc-seconds or better, equivalent to the Hubble Space Telescope's performance at visible wavelengths. This will allow astronomers to see features just a few AU across in nearby planetary systems, and might be enough to observe the effects of young planets ploughing through their accretion disks. The array could begin taking readings in 2007.

Once data from these arrays start to flow, our understanding of protoplanetary disks is likely to be transformed. Together with the new planet-detection techniques, which are poised to reveal the true diversity of extrasolar worlds, they should provide the data to eliminate some theories, and constrain others. “We're getting the census done right now,” says Boss. “In another ten years we'll really know what's out there.” ■

Dan Falk is a freelance science writer in Toronto.

1. Cody, A. M. & Sasselov, D. D. *Astrophys. J.* **569**, 451–458 (2002).
2. Mayer, L., Quinn, T., Wadsley, J. & Stadel, J. *Science* **298**, 1756–1759 (2002).
3. Trilling, D. E. et al. *Astron. Astrophys.* **394**, 241–251 (2002).

The Extrasolar Planets Encyclopaedia

♦ www.obspm.fr/planets

Sub-Millimeter Array

♦ <http://sma-www.harvard.edu>

Combined Array for Research in Millimeter-wave Astronomy

♦ www.mmarray.org

Atacama Large Millimeter Array

♦ www.eso.org/projects/alma

♦ www.alma.nrao.edu

Progressing towards a biological names register

How taxonomy could harness the indexing and organizational powers of the Internet.

Sir — There is much current debate about the decline in the number of taxonomists at a time of increasing need to monitor and manage our biodiversity resources. Various people and organizations have called for a reinvention of taxonomy and for new strategies in the delivery of knowledge, with an emphasis on bio-informatics solutions using the Internet (see, for example, H. C. J. Godfray *Nature* **417**, 17–19; D. Agosti and N. F. Johnson *Nature* **417**, 222; 2002).

Specialist aggregators (such as the Australian Biodiversity Information Facility, AlgaeBase, Antbase and FishBase) are forming unified global biodiversity catalogues for use by general aggregators such as Species 2000 and the Integrated Taxonomic Information System, or by more general web-based biodiversity projects such as the Tree of Life. Some Internet resources integrate information from distributed sources (for example, the Species Analyst, Fungalweb, micro*scope, Australia's Virtual Herbarium, and Global Searcher). What is missing is a unifying device to draw all these together on the Internet, and the tools to use the indexing and organizational power of taxonomy.

Virtually every biological database uses names, so these have the potential to act as a unifying device for biodiversity bioinformatics. This has not happened because — despite the best efforts of the codes of nomenclature — organisms may have more than one name, and names change in response to taxonomic insights. The situation can be managed by taxonomic name servers that map alternative names against each other.

Name servers can act as a biodiversity thesaurus, reconciling alternative names for the same taxon. Reconciliation functions can be embedded in Internet search engines or website search functions to allow novices or machines to find information under all of the relevant names, not just the one used to initiate a search. At a grander level, name servers may 'read' the name fields of biological databases, match entries against a master names list, and replace or add a standard entry drawn from a universal names register. Once populated with names, name servers offer realistic prospects for low-cost, accessible warehousing of distributed biological data, and of machine-to-machine dialogue about biology.

Conceptually, biological name servers are an extension of database thesauri or traditional synonymy lists, which catalogue alternative names. Software systems capable

of managing names are not common. Platypus, a database package for taxonomists developed through the Australian Biological Resources Study, is a stand-alone computer-based software, but systems capable of working through the Internet, such as that used by the Biodiversity Conservation Information System (BCIS), have considerably more appeal because they can index and organize distributed information. The Taxonomic Name Server (TNS) — part of uBio, the Universal Biological Indexer and Organizer project (www.ubio.org) — and a new concept called Octopus are emerging as broad, Internet-based biological name servers. TNS is the most advanced, with a development version in use by micro*scope (www.mbl.edu/microscope).

There are probably about 10 million names for about 1.7 million species. Name servers must deal with synonyms (alternative names for the same entity) and homonyms (identical names for different entities). Name servers will acquire their content from aggregators, the public domain, taxonomists and the literature. They will share content with each other, and return value-added information to the public domain. At this time, aggregators probably hold only about 10% of required names, and the rate of addition is slow — comprehensive lists are said to be 10–25 years away. This creates a challenge to accelerate and coordinate the compilation.

Currently, nomenclatural information about species is assembled by alpha taxonomists and passes through taxonomic revisions en route to expert aggregators such as FishBase or AlgaeBase, and then to general aggregators such as Species 2000. The rate-limiting elements are availability of taxonomic experts and the intellectual

effort of compiling definitive information.

A different approach is required if we intend to make rapid progress in the near future. Compiling names of organisms within an Internet-based name server will evade the rate-limiting step, allowing definitive taxonomic information to be acquired from authoritative remote sites as it becomes available. Using this approach, the uBio project is developing a universal indexing system for biology and expects to deliver a comprehensive and unified list of genera early next year. Once this list is in place within a homonym-aware environment, taxonomic service providers can read large lists of species names from, for example, the National Center for Biotechnology Information or BIOSIS, and assign each species to its correct location very quickly.

A major agency is needed to coordinate such a development, and to put in place an expert review panel (of at least 200 taxonomists) to oversee continuing taxonomic input. The Global Biodiversity Information Facility (GBIF) is the most obvious contender. Such an agency can coordinate the process, standardize products and processes, establish the credentials of the peer-review process, seek financial support for the peer-review panel, represent the interests of the compilation to governments and intergovernmental bodies, and ensure its integration with other initiatives. Most significantly, the involvement of GBIF will help to ensure that the names register and associated services remain in the public domain, and that all individuals who contribute are given appropriate recognition.

David J. Patterson

School of Biological Sciences, Heydon Laurence Building, University of Sydney, Sydney, New South Wales 2006, Australia

Peer review and the rewards of open access

Sir — Eugene Koonin's suggestion in Correspondence (*Nature* **422**, 374; 2003) that a journal should reward swift peer-reviewers by equally swiftly processing any articles that they themselves submit is an interesting one. Perhaps the converse should apply too: submissions from those who persistently underperform as reviewers should have their own work reviewed on a slower timescale.

An alternative reward system for good reviewers can be used in the open-access model of publishing, in which authors of

accepted articles pay a processing charge so that there are no subscription or access charges to the journal. In this model, swift reviewers can be rewarded by a reduction in this charge. A twist is applied by some journals published by Berkeley Electronic Press (www.bepress.com), in which submitting authors do not have to pay the processing charge if they contract to provide a timely review of an agreed number of papers. But they do have to provide credit-card details and are subsequently charged if they fail to deliver their reviews on time.

Peter Newmark

Editorial Director, BioMed Central, 34–42 Cleveland Street, London W1T 4LB, UK

Out, out, brief candle?

Time is slowly running out for life on Earth.

Impossible Extinction: Natural Catastrophes and the Supremacy of the Microbial World

by Charles C. Cockell

Cambridge University Press: 2003. 192 pp.
£18.95, \$28

The Life and Death of Planet Earth: How the New Science of Astrobiology Charts the Ultimate Fate of our World

by Peter D. Ward & Donald Brownlee

Piatkus Books/Time Books: 2003. 256 pp.
£16.99/\$25

Norman H. Sleep

One of the great intellectual achievements of the past millennium was realizing that the Sun is a star and the Earth is a planet. Over 400 years, this idea has gone from an excuse to burn the astronomer Giordano Bruno at the stake to a banality reserved for trick questions such as: "What is the nearest star to the Earth?" Reaction to the rank odour of biblical catastrophism and creationism has prejudiced serious consideration of extra-terrestrial affairs in geology and biology. This has now changed, the main wake-up call being the discovery that an asteroid impact wiped out a vast number of species, including the dinosaurs, 65 million years ago.

A chief beneficiary of this change is astrobiology, the science of planetary habitability, including the effects of extraterrestrial processes on the evolution of life on Earth, which has become a hot topic. Charles Cockell, Peter Ward and Donald Brownlee add to the list of popular books on this subject. Both books focus on global sterilization and the finite tenure of life on a habitable planet, but Cockell concentrates on microbes, whereas Ward and Brownlee look at animals.

These treatments bring different emphases: that it is quite hard to sterilize all the microbes on a planet, and that the days of animals are numbered by a few hundred million years. But the authors agree on the salient conclusions of astrophysics, geology and biology. The Solar System formed about 4.5 billion years ago. Microbes inhabited Earth by the time of the first good geological record, 3.5–3.8 billion years ago, perhaps even earlier. Our Sun's useful life is limited: its luminosity has already increased by 30%, and will continue to increase, until, 7 billion years from now, the red-giant stage will finally incinerate the Earth. In addition, the tectonic processes that continually renew the Earth's surface are on the wane.

I found Ward and Brownlee's book the easier to read. Its organization around the



Flaming Dune: as the Sun heats up, the Earth could become increasingly hot, dry and inhospitable.

various vicissitudes that await life works well, using an analogy to the progressive failure of various organ systems of an aged person. We can expect that as the Sun waxes, the weathering of rocks will reduce atmospheric CO₂ to low levels, shutting down plant life and leaving little food or oxygen for animals. Still later, the greenhouse will become hotter, ending with either the loss of the ocean's hydrogen to space or a runaway period of oven-like conditions at the surface. The authors correctly note that the former possibility leaves us with Dune—a hot, dry planet that is still habitable by microbes, and perhaps by animals near the poles—whereas the latter leaves the surface sterile. In the meantime, the continents may group together, fouling the global climate, or plate tectonics may shut down, leaving us, as in the film *Waterworld*, with a water-covered planet from which all topography has been eroded.

In contrast, I found the organization of Cockell's book—on a trip of the Solar System once around the galaxy—rather contrived. As he admits, galactic effects on the Earth have been, and probably will remain, small. Any effect has more to do with our position relative to the galactic plane and to the spiral arms with giant molecular clouds and young stars than to absolute position relative to distant galaxies. This quibble aside, the book is worth reading. A highlight is the discussion of why we should not generalize too much from the

actual state of the Earth. Cockell focuses on the presence of our large Moon. I might add that intelligent deep-sea organisms could probably come up with many reasons why intelligence could not arise on land, and that organisms circling a star ejected from its galaxy might think the plane of the galaxy to be an unattractive place.

Both books have the obligatory and entertaining discussion about how a high-tech civilization could avoid or at least delay its doom. We all like to feel that our distant descendants, or at least our more distant microbial relatives, will survive. It is easy to come up with suggestions to be filed away for 100 million years. Importantly, both books show that the Earth is not a truly safe place even now. There are hazards that we can do something about: global warming by the burning of fossil fuels; the more distant threat of advancing glaciers; and the ever-present danger of impacts from asteroids and comets. Science provides both the technology to cause trouble for ourselves and the knowledge to foresee and avoid that trouble.

Astrobiology derives from several fields that until recently had little interaction. An inevitable result is that it is difficult for one or even two authors to cover fully this rapidly evolving field. Specialists will quibble with various details and prefer more treatment of their own fields; my own include the effects of global tectonics, geochemical cycles and large asteroid impacts on early life. More

D. MUENCH/CORBIS

bothersome is the fact that both books contain some errors, so care is needed in using them as secondary sources. For example, Ward and Brownlee give the formation rate of continents as 650–1,300 cubic kilometres per year, a factor of about 1,000 too high. And it seems unlikely that global biomass drops by several thousandfold in an ice age. In the other book, Cockell's Earth has a mainly iron mantle, and his surface of Venus is younger than 250 million years and still active, rather than all about 500 million years old.

Having said that, I recommend both books to the lay reader. They can also be quickly read by a professional astrobiologist and do provide insight. And they would serve as supplemental reading (but not as primary texts) for a non-major class on planetary habitability or astrobiology. ■

Norman H. Sleep is in the Department of Geophysics, Stanford University, Stanford, California 94305, USA.

A stratospheric success

Protecting the Ozone Layer: Science and Strategy

by Edward A. Parson

Oxford University Press: 2003. 396 pp. \$65

Protecting the Ozone Layer: The United Nations History

by S. O. Anderson and K. M. Sarma

Earthscan: 2002. 544 pp. \$65, £40

Charles Herrick

When a scientific finding emerges from the field or laboratory, it undergoes a profound transition, changing from a well-formulated, discipline-bound construct into a dynamic and perhaps unruly component of a complex and interactive social system. The societal value of science and technology is created at the interface between laboratory and society, residing in neither alone, to paraphrase my friend Daniel Sarewitz. There is no direct and necessary linkage between the structure of physical reality and the manner in which society applies its scientific understanding of that structure. Rather, the social application of scientific knowledge and technological achievement is mediated through the formation and execution of policies.

Two new books, both entitled *Protecting the Ozone Layer*, chronicle the stratospheric phenomenon of ozone depletion as a window to explore the complex and dynamic relationships between science, technology and public policy. Although no small chore, as they have a combined length of almost 1,000 pages, I encourage readers to tackle both books, as their complementarity is fortuitous and sometimes quite illuminating.

Drawing on such fields as international-

relations theory and political sociology, Edward Parson outlines the formulation and implementation of the Montreal Protocol on ozone-depleting substances and its subsequent annexes, in terms of constructs such as epistemic communities and transnational issue networks. This provides the theoretical basis for his account of how scientific assessments, as distinct from scientific results per se, contribute to the development and maturation of adaptive policy regimes. Anderson and Sarma, on the other hand, offer a historical account based primarily on documentation and the recollections of key players in the ozone-treaty process.

Parson weaves a convincing account of the dynamic conditions that caused the initial slow progression of ozone negotiations between 1978 and 1986, and that then underpinned the swift consolidation of the Montreal Protocol by 1988. Briefly, he argues that scientific findings, considered alone, have little influence on the process of policy formulation. Rather, scientific results acquire credibility, salience and legitimacy only when compiled and interpreted in the context of a formalized, multi-party assessment process. As he writes, assessments "make key scientific statements and their implications common knowledge among policy actors" in the sense "that all parties know them, all know that all know them, and so on". In this way, assessments serve to bound and delimit the domain of reasonable negotiation and, once established, hasten regime formation.

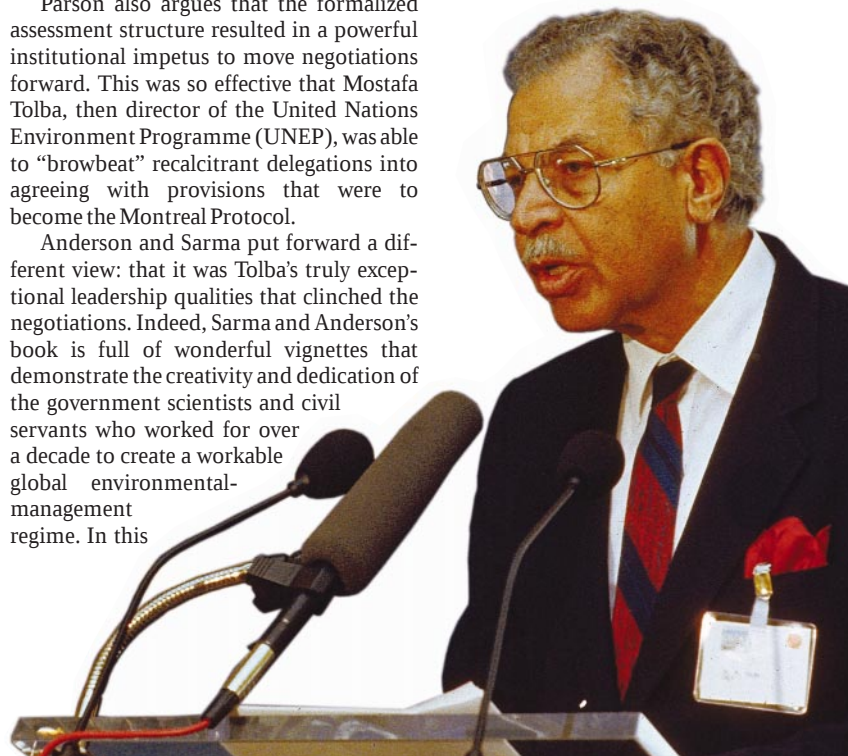
Parson also argues that the formalized assessment structure resulted in a powerful institutional impetus to move negotiations forward. This was so effective that Mostafa Tolba, then director of the United Nations Environment Programme (UNEP), was able to "browbeat" recalcitrant delegations into agreeing with provisions that were to become the Montreal Protocol.

Anderson and Sarma put forward a different view: that it was Tolba's truly exceptional leadership qualities that clinched the negotiations. Indeed, Sarma and Anderson's book is full of wonderful vignettes that demonstrate the creativity and dedication of the government scientists and civil servants who worked for over a decade to create a workable global environmental-management regime. In this

context, many of the individuals quoted or paraphrased cite the importance of informal groups and of a strong and inclusive consensus process that tended to override legalistic hair-splitting. So, did the institutional setting empower the actors, or did an exceptional group of actors combine to stretch the bounds of game-theoretic interpretation?

Both books do a good job of describing the evolution of stratospheric science and its relation to ozone depletion, chronicling the accomplishments of Sidney Chapman, G. M. B. Dobson, Lester Machta, Mario Molina and Sherwood Rowland, along with other pioneers of atmospheric science. They also describe how various stakeholder communities drew from this science base to support their positions and negotiations. I found myself wondering whether history would have been significantly different if industry had chosen an alternative field of battle and engaged the ozone-depletion debate from the perspective of cancer epidemiology, rather than atmospheric chemistry. Epidemiological studies show a rising incidence of melanoma, but evidence linking these skin cancers to increasing ultraviolet radiation, as the ozone layer was depleted, is complex and equivocal, and might have made a far less compelling case.

Neither book focuses heavily on the predicted and/or postulated effects of ozone depletion, merely noting, as Parson put it, that the policy debate was "pervaded by a general sense that effects did not matter".



Tough talking: effective scientific backing helped Mostafa Tolba forge the basis of the Montreal Protocol.

P. DEJONG/AP

How and why were these effects kept off the table? This seems to be a missed opportunity to tell a potentially powerful story about the socio-political construction of scientific assessments.

Like all good writing, these books leave the reader wishing for more: For example, I would have liked the authors to pick apart and compare the story of ozone diplomacy with that of other complex, science-based environmental-management issues, such as the various UNEP regional seas agreements, the Bonn Convention for the protection of the Rhine, or the Convention on International Trade in Endangered Species of Wild Fauna and Flora. Although not global in scope, these and other environmental regimes are also hypercomplex amalgams of scientific information, assessment process and political deliberation. Did ozone diplomacy follow well-established models of behaviour, break completely with past exercises in regime formation, or do a little of both?

Both versions of *Protecting the Ozone Layer* demonstrate that the path between scientific findings and policy formulation is anything but transparent, direct or linear. Instead, the relationship is mediated through a complex web of social, institutional and individual factors, acquiring context and meaning as they are embedded in larger narratives. ■

Charles Herrick is at Stratus Consulting Inc., 1920 L Street, Washington, DC 20036, USA.

Getting up to speed

Principles of Animal Locomotion

by R. McNeill Alexander

Princeton University Press: 2003. 376 pp.

\$49.50, £35

Stephen Gatesy

How did the chicken cross the road? A neurobiologist might answer that commands from the central nervous system, modulated by sensory feedback, generated cyclical walking movements. Alternatively, a physiologist could say that walking came about through myosin's interaction with actin, which created muscular tension to produce the necessary joint torques. On the other hand, a biomechanist might view the whole bird as a point mass rising and falling with each step, transforming gravitational potential energy into forward kinetic energy and back again, like an inverted pendulum. Such disparate responses highlight the wonderfully intricate and dynamic interplay that transpires when animals move through their environment.

In *Principles of Animal Locomotion*, R. McNeill Alexander approaches movement from a physical point of view. He seeks to explore the mechanical principles at work in organisms as different as burrowing earthworms and hovering hummingbirds. More-



How did the chicken cross the road?

over, he tries to account for the metabolic cost, and thus the relative energetic merit, of alternative forms of locomotion. These might seem like lofty goals, but Alexander is up to the task. His publications over the past six decades span such topics as swimming, scaling, gaits, elastic storage, and optimization (to name just a few), giving him the breadth and depth needed to summarize the field.

The book begins with a lucid summary of performance objectives — speed, acceleration, manoeuvrability, endurance, economy and stability — and their potential relationship to fitness. This is followed by a brief discussion of compromises (not all of the objectives are compatible), constraints (not all changes are possible) and optimization (which solution is best?). Alexander then covers muscle mechanics, energetics, scaling and recording techniques, before delving

into specific forms of locomotion. Terrestrial progression is addressed first through simple theoretical models and then by examining walking, running, hopping, climbing, jumping, crawling and burrowing in real animals. Aerial and aquatic locomotion are then tackled, although cilia-driven propulsion is not included. There is also a chapter on the effectiveness of devices for augmenting human locomotion.

Although this book is aimed at an audience ranging from advanced undergraduates to researchers and university teachers, I think that most undergraduates will be overwhelmed and intimidated if it is used as a textbook. Alexander packs in so much information that clarity is often sacrificed for coverage. His curt style is so direct and quantitative (equations are ubiquitous, starting on page 2) that every paragraph is extremely dense. But despite its failings as a textbook, it is an excellent reference tome. Citations (up to 2001) and 33 pages of references (of which 77 are authored by Alexander) will quickly direct readers to the relevant literature. I will use *Principles of Animal Locomotion* and recommend it to my graduate students.

Alexander's facility with simple, quantitative models is both a strength and a weakness. Regrettably, the translation from a living, breathing animal to a mathematical representation of its underlying physics is never explained. In creating such models, which details can be omitted, and which need to be retained? If Alexander wants to appeal to a broad range of scientists, he must take

Standing the test of time

The centenary volume of Munich's Deutsches Museum, *Ingenious Inventions and Masterpieces of Science and Technology*, is a celebration of the long and often troubled history of one of the world's oldest, and largest, technology museums.

The Deutsches Museum was founded in 1903 by Oskar von Miller, a pioneer of electrical engineering. Work on its permanent home — an architectural jewel on an island in the Isar river, which flows through Munich — was begun in 1906, but its completion was delayed first by war and then by hyper-inflation. The building was finally opened in 1925 with great fanfare in a celebration that is sometimes referred to as the Weimar Republic's last party.

A few happy years followed before history struck again. The Nazis didn't like von Miller's cosmopolitan aspirations for the museum, and the Allied forces' bombs had no pity, destroying much of the building as well as around 20,000 display items. But the museum survived and remains one of the world's most significant technology museums.

This coffee-table book, edited by the museum's director, Wolf Peter Fehlhammer, includes illustrations of treasures from the past along with current exhibitions. These range from



Otto von Guericke's first vacuum pump and hemispheres, forged in 1663, to the 1962 recreation of the Altamira cave with its Stone Age paintings, to the Helios space probes, solar observatories constructed in Munich and launched in the 1970s. The illustration shows a timepiece dating from 1630 from southern Germany, one of the earliest instruments for measuring time.

Science in culture

Nets and neuroses

The obsessional art of Yayoi Kusama uses repetition to express her fears of obliteration in infinity.

Martin Kemp

It is increasingly common for artists to make a conscious effort to embed science into their creative procedures. We accordingly feel justified in analysing their works in relation to the relevant sciences. But what about works that evoke strong scientific resonance without any evidence that their creators were thinking about science at all? A striking case in point are the huge paintings and installations by the Japanese artist Yayoi Kusama.

This dilemma is encapsulated by the reaction of Philip Campbell, editor of *Nature*, when he visited the retrospective of Kusama's works at the Bass Museum of Art in Miami Beach, Florida, housed in a new gallery designed by Arata Isozaki. Campbell says that "her flat patterns, such as the infinity nets, are fascinating and seem to have an elusive natural character, as opposed to an artificial character". He adds: "It's hard to judge the power of these sometimes gigantic canvases." But Campbell suggests some themes that spring to mind, such as pattern analysis, and says they are reminiscent of crystallographic and biological patterns. Yet there is no sign in any of Kusama's many pronouncements of her conscious engagement with such things.

The 'infinity nets' are composed of cellular structures of potentially unlimited extension. Each cell is similar, leading to endless repetition, yet is unique. At one level the nets are "mechanical" and "empty", as Kusama says, yet they are no more empty and undifferentiated than cosmic space, plastic foam or cross-sections of cork.

One of the more recent variations on the theme is *Infinity Stars* from 1995, which stretches to 17 feet in length. This work exemplifies how the repeated shapes — in this case a myriad of round 'lights' punctuating a cellular membrane — have organized themselves around certain intuitive principles of distribution, packing and symmetry. The very act of repeatedly stringing out the reticulate structure sets parameters on its possible morphology, much like the physico-chemical constraints imposed on cells in tissues or the space-time dimensions of cosmic branes.

But it would be wrong to write about the infinity



Spot the difference: this detail from *Infinity Stars* shows how Yayoi Kusama uses repetition.

nets as if they are only formal exercises that exhibit an unconscious *rapprochement* with certain classes of scientific image. They arise instead from deep psychic motivations, and are saturated with Kusama's fears, obsessions and hallucinations. She has, since childhood, suffered from obsessive neurosis, and for the past 20 years has voluntarily resided in a mental hospital in Japan. She has been obsessed by the hypnotic effects of endless repetition and accumulation, and both proclaims and fears the obliteration of the individual in scaleless infinity and limitless time. She exploits repetition in a manner akin to chanting in religious rituals or to the mesmerizing quality of an endlessly repeated phrase that loses all focused content.

Kusama has been one of the most remarkable and fertile artists of the second half of the twentieth century. She arrived in New York as a 29-year-old in 1958 and contributed energetically to the heady

art world of the 1960s and 1970s. Andy Warhol's notorious Factory set the tone of social ferment and creative anarchy. Kusama herself participated uninhibitedly, staging outrageous happenings with nude performers. She also made large accumulative sculptures densely covered in phalluses, set up extravagant installations with mirrors, designed extreme fashion and wrote unsettling poetry.

There is no simple answer to our dilemma of whether Kusama's nets have any connection with science. At the level of conscious address, the answer is probably no. And in as much as their expressionistic role is to evoke subjective feelings of obliteration, depersonalization and hallucination, the answer is again no (or apparently so). However, if we look at how they make their effect on a viewer, by tapping into ubiquitous modes of repetitive pattern formation — the effect to which Campbell reacted — we can see that they can be powerfully resonant for scientific observers who delve into the order of things of a kind, albeit for their own, very different purposes.

Martin Kemp is in the Department of the History of Art, University of Oxford, Oxford OX1 2BE, UK.

Yayoi Kusama's work will be on show at the Bass Museum of Art in Miami Beach, Florida, until 11 May.



Yayoi Kusama: influential for almost 50 years.

readers step-by-step through the transition to a mechanical perspective, rather than just diving in. Most biologists, even those with training in mathematics and physics, don't think like engineers.

Finally, I lament the book's lack of phylogenetic foundation. Despite the first chapter's evolutionary tone, Alexander never returns to ancestry or history; instead, each organism is treated as a separate case study of adaptation. In a disappointingly short epilogue (just five-and-a-half pages), he attempts "some generalizations about locomotion",

but comes to few conclusions. His search for generalizations fails because he treats each organism as an independent point on a graph, rather than as a member of a hierarchical tree of life. From an evolutionary perspective, physical mechanisms that are shared among taxa are either homologous or convergent, not just common.

If most walking animals, whether vertebrate or arthropod, use an inverted pendulum mechanism to save energy, how many times has this evolved? How did organisms that were optimized for swimming evolve into

organisms optimized for running? And then how did runners evolve into organisms that are optimized for flying? Such transitions are completely ignored in this book.

Principles of Animal Locomotion is a valuable reference book written by a leader in the field. But it also serves as proof that enough studies have accumulated to warrant an evolutionary analysis of locomotor mechanics — a synthesis that I await eagerly. ■

Stephen Gatesy is in the Department of Ecology and Evolutionary Biology, Brown University, Providence, Rhode Island 02912, USA.

Puzzling out the past

Robert Hedges

We have been hearing a lot of the word 'verification' lately, and although in the international arena it has now been replaced by more primitive tests, it has set me thinking about how verification operates in my own field of archaeological science. As with other historical sciences — such as cosmology, geology or evolutionary biology — our methods are based on well-understood physics and chemistry; but our evidence of what happened is fragmentary, and what remains may be altered out of all recognition.

Archaeology suffers further difficulties in its reconstruction of the past, for here human behaviour is central and so we must engage, somehow, with the mental world of our forebears. As human mentality can encompass the most sophisticated acts (of deception, for example), it is not always satisfactory to rely on the present to explain the past, or to attempt to interpret behavioural evidence in a purely rational way. In short, we need to be able to verify the methods by which we collect and interpret evidence.

When working on a jigsaw or crossword puzzle, we can be confident that the piece, or word, is correct if it gives both a precise 'fit' in the local space and is consistent with the emerging global pattern. This view also applies to verification in research, highlighting the latent tautology of fitting pieces together mainly on the basis of their global sense.

This danger has been demonstrated more than once in the dating of archaeological events, especially when using a new method. The history of radiocarbon dating provides an interesting perspective. This technique measures the time elapsed since the death of an organism; that is, since it lost its metabolic connection with the source of radioactive carbon — the atmosphere. But such a date must still be connected to human activity.

For example, carved ivory excavated from an Anglo-Saxon site could turn out to be from much older mammoth bone; subtler differences in age or usage may not be so obvious.

The precision of the local 'fit' is also crucial, and this includes wholesale assumptions, the past validity of which cannot be checked directly. Verification of radiocarbon dates is possible, however, on materials of absolutely known age, such as large pieces of wood in which annual rings can be counted. In such instances, consistent discrepancies have been found, and as well as allowing radiocarbon dating to be calibrated, these examples are now seen as testament to changes in the solar magnetic field and in terrestrial carbon cycles (such as ocean circulation).

The problems of verification become more acute when dating events that are so far in the past (beyond 50,000 years) that all of the radioactive carbon has decayed. Many of the methods available depend on the past environment, such as the accumulating effects of local radioactivity, but are forced to assume a homogeneous history for the sample. Whenever possible, corroboration through the use of multiple dating methods gives much-needed confidence to the 'fit'. For example, the dating of Neanderthal flints in Kebara Cave, Israel, by thermoluminescence gave similar ages to those estimated by electron spin resonance for animal teeth in the same strata. In another instance, controversy over the proposed 1.9-million-year age of the KBS volcanic tuff in Kenya was resolved when results from potassium-argon and fission-track methods were compared.

The bones of our ancestors, like our own, are made from the very atoms that they ate. In particular, they record the isotopic composition of the diet, albeit with some metabolic twists, and from this we are learning to infer, very roughly, the contribution of marine food, and of animal versus plant protein, to the diet. Much of this is founded on laboratory animal models, but estimates can

Archaeological verification

In piecing together our ancestors' habits, corroborating lines of evidence are vital if we are to be confident that the emerging jigsaw reveals the true picture.

be verified by examining living populations, even though the promiscuity of modern dietary habits, and the problems of sampling, make this an imperfect solution.

My colleagues and I recently had an opportunity to deliver a 'two-dimensional fit' in the verification of evidence on how our ancestors fed themselves. Between Serbia and Romania, the River Danube flows through a limestone gorge, and on its banks are archaeological sites of human settlements. How important were freshwater fish to people in this region? The scattering of fish bones is only a clue. The human bones indeed contain an increased abundance of the heavy isotope of nitrogen, as would be expected for aquatic food webs, but this is not evidence enough to be sure of a fishy diet. Is there another clue that will corroborate the hypothesis?

The clincher comes from radiocarbon dating. The site contains human burials with, in several cases, spear points of animal bone still fatally embedded in the skeleton. However, the dates measured for the humans are apparently several centuries older than those for the spears. This discrepancy is neatly accounted for if they consumed a substantial quantity of fish. This is because, as the Danube flows through the gorge, it dissolves geological-aged limestone, whose (very old) carbonate carbon is fixed and passed up through the aquatic food chain to human bone collagen. The bones of fish-eaters would therefore contain less radioactive carbon than those of the animals from which the spears were made — who would have grazed on terrestrial carbon sources.

Rarely can we be as confident as with the double fit provided here. Answering these questions usually requires patience and discipline to ensure that bringing the past to light is not a matter of two shaky steps forward and one slipped step back. ■

Robert Hedges is at the Research Laboratory for Archaeology, University of Oxford, 6 Keble Road, Oxford OX1 3QJ, UK.

FURTHER READING

Ferguson, C. F. *Science* **159**, 839–846 (1968).
DeNiro, M. J. *Am. Sci.* **75**, 182–191 (1987).
Cook, G. T. *et al. Antiquity* **76**, 77–85 (2002).



CORBIS

Wise fathers

John D. Reynolds and Ben C. Sheldon

As Shakespeare put it, "It is a wise father that knows his own child". Male bluegill sunfish do: they adjust their behaviour towards their young according to how sure they are of being the real father.

Studies of humans suggest that maternal relatives are more likely to comment on a newborn baby's resemblance to its putative father than to its mother¹. Perhaps these comments provide reassurance about a father's likelihood of being the true father of the child. This interpretation makes evolutionary sense if, as theory predicts, males adjust their level of parental care to their certainty of paternity. Although this expectation fits with the widespread use of DNA tests in messy divorce cases, it has proved extremely difficult to determine whether males in other species play by similar rules.

An elegant new study by Bryan Neff, on page 716 of this issue², breaks through the usual practical difficulties and shows that male bluegill sunfish (*Lepomis macrochirus*; Fig. 1) do indeed adjust their parental behaviour in response to their certainty of paternity. These findings advance our understanding of the evolution of parental care and raise new questions about the conditions under which offspring should reveal their identity to parents, or conceal it.

The study of bluegill sunfish overcomes a stumbling-block that has hampered many previous attempts to test the theory of parental investment in relation to paternity^{3,4}: how to manipulate certainty of paternity. 'Certainty' is not something that can be measured directly. Many researchers have substituted actual measures of paternity using genetic markers³. The hope is that, although the study species won't be able to read DNA bands on gels, it will have picked up some correlate of its paternity, such as the behaviour of its mate towards other suitors. A much more satisfactory approach would be to manipulate the actual cues that males are known to use when assessing their paternity. This is where Neff's study of bluegill sunfish comes in.

Bluegill sunfish are native to most of the United States and adjacent Canada and Mexico, where they nest in colonies in lakes (Fig. 1). Intense competition among males during the breeding season has led to the evolution of two distinct life-history pathways. Males termed 'parentals' defend nest sites, attract females, and then care for the eggs and newly hatched offspring. The others mature at an earlier age as 'cuckolders' and steal fertilizations from parentals either by darting into nests at the critical moment of spawning ('sneakers') or by mimicking



Figure 1 A colony of breeding bluegill sunfish in Lake Opinicon, Ontario, Canada. Parental males are tending their nests, while females ready to spawn are swimming higher up in the water. Neff² has shown that male bluegills use two cues to assess how likely it is that they are the father of the offspring: the more certain they are of their paternity, the more attentive they are as parents.

females, apparently fooling the parental male into thinking he has attracted two females at once⁵. Sneakers are particularly effective, fertilizing 89% of the eggs released by a female during the 8% of spawnings in which they participate⁶.

The occurrence of sneaking thus provides a cue that parental males could use as a guide to their paternity. The second cue is more surprising, but has been confirmed by controlled experiments⁷. Parental male sunfish can apparently assess their relatedness to newly hatched fry using water-borne odour cues; the mechanism is unknown, but other studies of fish suggest a role for genes in the major histocompatibility complex (MHC) in olfactory discrimination of kin⁸ and potential mates⁹.

Neff exploited these mechanisms of assessing paternity in two experiments that examined the males' willingness to defend the nests against an egg predator. In the first, parental males in the midst of spawning were exposed visually to four sneaker males, enclosed in transparent plastic containers so that they could not fertilize any eggs. Control males were exposed to empty containers. As predicted, males reduced their level of care

during the egg phase when they were tricked into expecting lower paternity. Then, male care was tested a second time after the eggs hatched, when the second mechanism for assessing paternity — olfaction — was predicted to restore the certainty of paternity of the experimental males. This is exactly what happened: these males increased their care when the new information suggested that their paternity was not lower than in the control group. These reductions and increases in care fit perfectly with predictions based on the two mechanisms of assessing parentage.

In a second experiment, Neff transferred one-third of a clutch of eggs between nests of parental males, and, as before, assessed the nest defence of parental males before and after hatching. This experiment could therefore only influence male behaviour through the second mechanism for paternity assessment (olfaction), as males seem to be unable to distinguish their relatedness to offspring before hatching occurs⁷. As expected, there was no difference in the behaviour of experimental and control males before egg hatching, but the experimental males decreased their intensity of defence after the eggs hatched.

Two aspects of these experiments are

MARTIN R. GROSS

particularly notable. First, the changes in behaviour were measured for individual males, which is likely to have reduced the influence of other sources of variation between individuals. Second, the responses to the two experiments are in different directions, showing that males adjust their level of care both up and down in response to changes in certainty of paternity. Taken as a whole, these experiments provide convincing evidence that male bluegills adjust their behaviour in response to their certainty of paternity.

This research suggests some fascinating areas for future work. It would be interesting to follow the fortunes of males through successive spawning bouts to see whether their decisions to adjust care in relation to paternity enhance their lifetime reproductive output, as predicted by life-history theory. The male's ability to determine his relatedness to offspring on the basis of odour cues raises the question of why offspring sired by cuckolders have not evolved an ability to conceal their identity, and whether they might use any tricks to exploit the males that guard them, in the same way that nestling cuckoos manipulate their foster parents¹⁰. Theory

to address the concealment of identity is already partly in place^{11,12}: bluegill sunfish might provide just the system to add empirical flesh to this framework.

John D. Reynolds is at the Centre for Ecology, Evolution and Conservation, School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, UK.

e-mail: reynolds@uea.ac.uk

Ben C. Sheldon is at the Edward Grey Institute, Department of Zoology, South Parks Road, University of Oxford, Oxford OX1 3PS, UK.

e-mail: ben.sheldon@zoo.ox.ac.uk

1. Daly, M. & Wilson, M. I. *Ethol. Sociobiol.* **3**, 69–78 (1982).
2. Neff, B. D. *Nature* **422**, 716–719 (2003).
3. Whittingham, L. A. & Dunn, P. O. *Curr. Ornithol.* **16**, 257–298 (2001).
4. Sheldon, B. C. *Phil. Trans. R. Soc. Lond. B* **357**, 341–350 (2002).
5. Gross, M. R. & Charnov, E. L. *Proc. Natl Acad. Sci. USA* **77**, 6937–6940 (1980).
6. Fu, P., Neff, B. D. & Gross, M. R. *Proc. R. Soc. Lond. B* **268**, 1105–1112 (2001).
7. Neff, B. D. & Sherman, P. W. *Anim. Cogn.* (in the press).
8. Olsen, K. H., Grahm, M. & Lohm, J. J. *Chem. Ecol.* **28**, 783–795 (2002).
9. Reusch, T. B. H., Haberli, M. A., Aeschlimann, P. B. & Milinski, M. *Nature* **414**, 300–302 (2001).
10. Davies, N. B. *Cuckoos, Cowbirds and Other Cheats* (Poyser, London, 2000).
11. Pagel, M. *Anim. Behav.* **53**, 973–981 (1997).
12. Johnstone, R. A. *Proc. R. Soc. Lond. B* **264**, 1547–1553 (1997).

Astronomy

Wrestling monsters in deep space

Lennox Cowie

Observing distant galaxies is always problematic. But when it comes to the biggest star-forming galaxies, far across the Universe, only indirect approaches can give astronomers any handle at all.

One of the big questions in observational cosmology is when and where the stars in galaxies formed. Giant optical telescopes can now pick out galaxies at huge distances across the Universe, but curiously they miss the galaxies in which the greatest number of stars are forming. Much of the star formation in the Universe occurs in bursts, and the processes that fire up these monster galaxies also produce huge amounts of dust that blot out starlight at most wavelengths, including optical ones. The energy that is absorbed by the dust is re-radiated at much longer wavelengths (in the submillimetre range), and observations at these wavelengths have only recently revealed the presence of these hidden systems.

The fact that so little optical light escapes from the giant star-formers also makes it very difficult to determine how far away they are — distance is usually derived from the spectra of optical emission from galaxies. Two new approaches to finding the distances, or redshifts, of monster galaxies are now reported: by Chapman *et al.*¹ on page 695 of this issue, and by Wiklind² in the *Astrophysical Journal*. Wiklind uses the spectral shape of the galaxies at long wavelengths

to estimate their redshifts. Chapman *et al.* use radio data to locate the galaxies accurately and then obtain the best possible optical spectra.

In 1996, the COBE satellite measured the energy density of the Universe at submillimetre wavelengths and made the remarkable discovery that, over the lifetime of the Universe, galaxies have radiated as much energy at submillimetre as at optical wavelengths³. Clearly there had to be an as-yet undetected population of galaxies that has produced this energy — which is remarkable because such a population would have to have formed as many stars as all the galaxies seen in optical observations. But this population does exist: the Submillimetre Common-User Bolometer Array, or SCUBA, on the 15-m James Clerk Maxwell Telescope on Mauna Kea, Hawaii⁴, has found a huge number of luminous galaxies^{5,6} radiating at submillimetre wavelengths — enough to produce nearly all of the submillimetre light seen by COBE⁷.

Simply imaging the galaxies, however, is not sufficient. To measure how they evolved with time, and to map the history of star formation, we also need to know the distances

to these galaxies. This has proved to be a hard problem. Identifying the optical counterparts to the submillimetre sources is difficult because observations at long wavelengths are limited in their positional accuracy, even when very large telescopes are used. Ultimately, this problem will be resolved with a new generation of submillimetre-telescope arrays, the first of which is now coming online on Mauna Kea⁸. But for the moment it is still difficult to match up the images of the same object seen at optical and at submillimetre wavelengths.

The first attempts to measure redshifts for the submillimetre sources involved identifying every one of the handful or so optical galaxies near the position of each detection. The method was tedious, but distances could be measured for about a quarter of the submillimetre sources and were typically about two-thirds of the way across the Universe⁹. A more promising alternative, however, is to locate the submillimetre counterparts through radio observations. This works because there is a relatively tight empirical correlation in local star-forming galaxies between radio emission and thermal emission from dust¹⁰. The locations of about 60% of the bright submillimetre sources can be pinpointed in this way¹¹.

Chapman *et al.*¹ have drawn on this result and used radio observations to locate the optical counterparts to the submillimetre sources. They then obtained optical spectra over very long exposure times to identify emission lines in the small amount of optical light that emerges from the galaxies. Chapman *et al.* have measured redshifts for ten submillimetre sources, which is a substantial increase in the number of such sources with known redshifts. Even so, this approach unfortunately only samples sources from the same bright submillimetre population that had previously been identified. Although these redshifts are obtained with much less effort, distances to most of the sources still cannot be measured.

In the not so distant future, it should be possible to measure redshifts directly using the Atacama Large Millimeter Array — an ensemble of 64 dishes, each 12 m in diameter, to be built in Chile. Until then, the distances to most submillimetre sources can only be measured using crude redshift estimators. The most widely used estimator is based on the ratio of the amount of emission, or flux, at submillimetre wavelengths to the flux at radio wavelengths¹². Unfortunately, this method fails when the submillimetre sources become too faint — too distant — to detect in the radio. Wiklind² shows that redshifts for these distant sources can be estimated using an alternative method that is based on the shape of the submillimetre region of their spectra. But even this method requires more sensitive submillimetre imaging observations than have yet been made

and will depend for its success on the next generation of submillimetre instruments.

Despite their limitations, these methods devised by Chapman *et al.*¹ and by Wiklind² take us towards a better understanding of star-forming systems, while the next generation of telescopes is awaited. But for the moment, many of the star-forming monsters of deep space will keep their secrets. ■

Lennox Cowie is at the Institute for Astronomy, University of Hawaii, 2680 Woodlawn Drive, Honolulu, Hawaii 96822, USA.
e-mail: cowie@ifa.hawaii.edu

Medicine

Collateral damage repaired

Lawrence Steinman

Multiple sclerosis is characterized by immunological attacks across a wide front in the brain and spinal cord. In mice, the damage can be partly repaired by neural precursor cells, delivered into the blood or spinal fluid.

Multiple sclerosis affects nearly one million people worldwide, subjecting people from young adulthood onwards to repeated immunological attacks on the brain and spinal cord. Twice as many women as men are afflicted with the disease. The effects vary depending on where exactly in the nervous system the attacks occur, but paralysis, blindness, loss of sensation and a lack of coordination are among the types of devastation wrought by an immune system gone awry. Until now, treatment strategies have generally been aimed at blocking the autoimmune attacks and reducing the amount of collateral damage caused. On page 688 of this issue, Pluchino and colleagues¹ describe a complementary approach — repairing some of the harm already done.

The immunological attacks that underlie multiple sclerosis damage the brain and spinal cord in several ways. First and foremost, the immune system mistakes myelin — the fatty sheath that insulates nerve cells — for foe instead of friend, and releases rounds of friendly fire. The myelin thereby becomes damaged and inflamed. That is also the fate of oligodendroglial cells, specialized types of non-neuronal (glial) cells in the brain, which both produce myelin and act as insulation in their own right. The inflammatory damage interferes with the flow of electrical impulses along underlying nerve fibres. The nerves themselves may also be harmed, ultimately leading to their demise. Finally, astrocytes — another type of glial cell — enlarge and proliferate in a manner analogous to the way in which scar tissue forms around injuries elsewhere in the body. This astrocytic scarring adds to the difficulty of propagating electrical activity down nerve fibres, by dispersing electrical impulses.

1. Chapman, S. C., Blain, A. W., Ivison, R. J. & Smail, I. R. *Nature* **422**, 695–698 (2003).
2. Wiklind, T. *Astrophys. J.* (in the press); preprint astro-ph/0301403 at <http://arXiv.org> (2003).
3. Puget, J.-L. *et al. Astron. Astrophys.* **308**, L5–L8 (1996).
4. Holland, W. S. *et al. Mon. Not. R. Astron. Soc.* **303**, 659–672 (1999).
5. Barger, A. J. *et al. Nature* **394**, 248–251 (1998).
6. Hughes, D. H. *et al. Nature* **394**, 241–247 (1998).
7. Cowie, L. L., Barger, A. J. & Kneib, J.-P. *Astron. J.* **123**, 2197–2205 (2002).
8. Moran, J. M. *Proc. SPIE* **3357**, 208–219 (1998).
9. Barger, A. J. *et al. Astron. J.* **117**, 2656–2665 (1999).
10. Condon, J. J. *Annu. Rev. Astron. Astrophys.* **30**, 575–611 (1992).
11. Barger, A. J., Cowie, L. L. & Richards, E. A. *Astron. J.* **119**, 2092–2109 (2000).
12. Carilli, C. L. & Yun, M. S. *Astrophys. J.* **513**, L13–L16 (1999).

Most existing treatments for multiple sclerosis aim to block the immunological attacks on a wide front^{2,3}. One method involves interfering with the adhesive, Velcro-like molecules that immune cells use to attach themselves to blood vessels as they prepare to move from the circulation to the brain. A decade ago, experiments⁴ in an animal model of multiple sclerosis, called experimental autoimmune encephalomyelitis (EAE), showed that blocking a key adhesion molecule — $\alpha 4$ integrin, found on the surface of attacking immune cells — could reverse paralysis. The results of clinical trials of patients with multiple sclerosis⁵

suggest that the same approach can block clinical attacks. Other strategies involve weakening rogue immune cells by reducing their production of inflammatory chemicals (including molecules such as tumour-necrosis factor- α , a type of cytokine protein), and by inhibiting destructive enzymes known as metalloproteases².

These approaches are still being tested, but all hold the promise of impeding future immune attacks on myelin and nerve fibres. Equally important, however, is the repair of existing damage, and this is where Pluchino *et al.*¹ come in. The authors started by isolating neural precursor cells from the lining of the brains of mice, and then, remarkably, they used them to attenuate paralysis and neurological dysfunction in EAE. Although the exact nature of these cells is debatable, they fulfil many of the definitions of adult neural stem cells. They were isolated from a region called the periventricular zone, located next to the internal canals of the brain — the ventricles — where spinal fluid nourishes the nervous system. They are also multipotent, which means that they can develop into various specialized cell types when provided with appropriate signals.

Surprisingly, Pluchino *et al.* found that — like the attacking immune cells — the neural precursor cells express $\alpha 4$ integrin. So, once injected into the blood or spinal fluid, they can move to points of inflammation within the brain and spinal cord of mice with EAE. There they somehow become involved in processes that decrease the levels of inflammatory chemicals (such as tumour-necrosis factor- α) and metalloproteases. The neural precursors also reduce the astrocytic scarring associated with brain inflammation. In

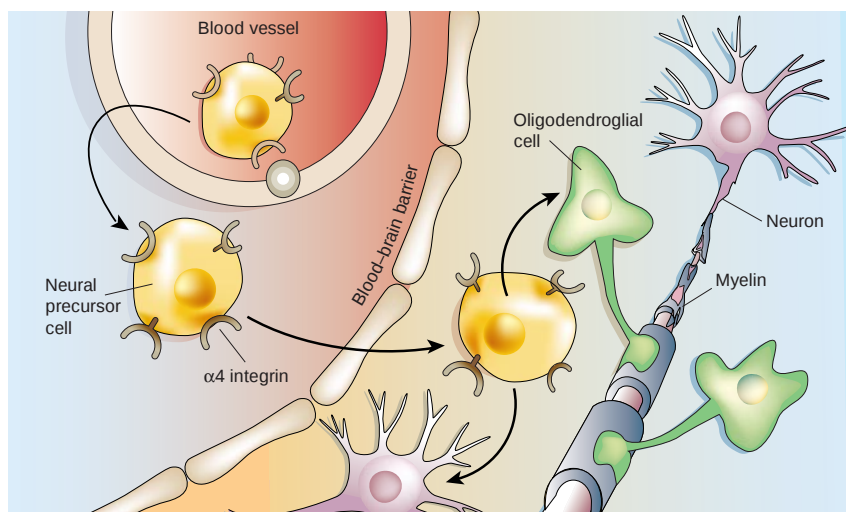


Figure 1 Rescuing the nervous system. In people with multiple sclerosis, the immune system attacks both nerves and myelin, the fatty sheath that encompasses nerve fibres. Pluchino *et al.*¹ have isolated neural precursor cells from the brains of mice, and injected them into the blood or spinal fluid of mice with experimental autoimmune encephalomyelitis — an experimental model of multiple sclerosis. The cells express the $\alpha 4$ integrin protein, and may use this to move from the blood into the brain. There the cells differentiate into both oligodendroglial cells, which generate myelin, and new neurons. The mice show a marked improvement in their symptoms.

damaged areas, they give rise to a pool of new myelin-producing cells (the oligodendroglial cells), and to new neurons (Fig. 1). They also produce growth factors, including ciliary neurotrophic factor, that may provide a restorative milieu⁶. Most importantly, the symptoms of mice improve even when the precursor cells are delivered to animals already suffering an attack of paralysis. Disability wanes, and electrical conductivity along nerve fibres increases.

The potential of strategies such as this to treat neurological damage on a wide front is impressive. Although some neurological disorders, such as Parkinson's and Huntington's diseases, are confined to specific brain regions, others, like multiple sclerosis and Alzheimer's disease, affect much broader areas. When the damage is confined, a local injection of neural precursors might be beneficial. Similarly, in previous studies^{7,8}, precursors of myelin-generating cells have been transplanted directly into demyelinated brain regions. But this would be an impractical means of coping with the widespread damage seen in multiple sclerosis, which must be tackled differently. Until now, this requirement seemed daunting, but the results of Pluchino *et al.* put matters in a new light, by showing that neural precursors can be injected into the blood or spinal fluid and still find their way to the many areas where they are needed. One point of particular interest here is that these cells hitch a ride into damaged sites by using $\alpha 4$ integrin—the very molecule that mobilizes the immunological attack^{2,4,5}.

To give such cell-based strategies the best possible chance, it will be imperative to reduce the risk that newly formed myelin-

producing cells will be targeted in another round of friendly fire^{2,9}. But on both fronts—in silencing the autoimmune attacks and in repairing the brain damage—there is, I believe, good reason to be optimistic. Many attractive methods for dampening the autoimmunity that is characteristic of multiple sclerosis are under development. These include broad-scale tolerization with myelin-derived peptides² and with genes encoding myelin proteins^{2,3,9}. They can perhaps be combined with well-known drugs such as statins, which have recently been shown¹⁰ to be extremely effective in suppressing autoimmunity. It should be feasible to stop collateral damage. And once the immune system has been made to surrender, the molecules at fault can perhaps be turned to help promote rehabilitation. If sufficient numbers of human neural precursor cells can be collected, and if we can work out how to make these cells proliferate and differentiate, then the results of Pluchino *et al.* might be translated into a treatment that eliminates collateral damage in multiple sclerosis. ■

Lawrence Steinman is in the Department of Neurological Sciences and Neurology, Interdepartmental Program in Immunology, Stanford University, Stanford, California 94305, USA. e-mail: steinman@stanford.edu

1. Pluchino, S. *et al.* *Nature* 422, 688–694 (2003).
2. Steinman, L. *Nature Immunol.* 2, 762–765 (2001).
3. Lock, C. *et al.* *Nature Med.* 8, 500–508 (2002).
4. Yednock, T. *et al.* *Nature* 356, 63–66 (1992).
5. Miller, D. *et al.* *New Engl. J. Med.* 348, 15–23 (2003).
6. Linker, R. *et al.* *Nature Med.* 8, 620–624 (2002).
7. Archer, D., Cudde, P., Lipsitz, D. & Duncan, I. *Nature Med.* 3, 54–59 (1997).
8. Imaizumi, T., Lankford, K., Burton, W., Fodor, W. & Kocsis, J. *Nature Biotechnol.* 18, 949–953 (2000).
9. Garren, H. *et al.* *Immunity* 15, 15–22 (2001).
10. Youssef, S. *et al.* *Nature* 420, 78–84 (2002).

bacteria (known collectively as rhizobia). Through analysis of these nodules, they could then separate and dissect the processes of nitrogen reduction, assimilation of ammonium into amino compounds, and transport between the two partners. To the authors' surprise, the host plant cells could not assimilate ammonium when they were nodulated by bacterial mutants in which amino-acid transport was blocked.

There were two aspects to this observation. First, mutants that could fix nitrogen at rates comparable to the wild-type bacteria could not pass the products on to the host cell unless they were supplied with an amino acid, probably glutamate, by the host cell. Second, the host cell could not assimilate ammonium from bacteria unless it was also supplied with another amino acid, aspartate. Lodwig *et al.* propose that these two transport systems may have distinct functions in symbiosis (see Fig. 4 on page 725). One serves to import glutamate from plant to bacteria, and the other to export aspartate from bacteria to plant. So each side can impose a sanction on the other, by withholding a vital amino acid. If this circuit is in place, bacteria can export ammonium and ensure both their own amino-acid supply and that of their host. Thus, both sides have a strong interest in maintaining the marriage.

Before a host plant accepts bacteria into this intimate association, an intricate dialogue occurs between the two partners, which tests their mutual compatibility. Events begin in the soil, when plants and rhizobia exchange signals. They proceed via 'infection pathways' and nodule development (Fig. 1 shows a variety of nodule types). And they culminate in the formation of symbiotic units such as those studied by Lodwig *et al.*¹. But does this courtship always end in harmony? Unfortunately not.

Problems may occur at any stage, and two are illustrated by the work of Lodwig *et al.* First, with bacterial mutants that can induce nodulation but cannot allow ammonium assimilation, numerous small, 'ineffective' nodules result, typical of those sometimes found in nature. In this case, host control over the number of nodules produced³ is depressed. Second, mutants that cannot effectively use the host products of photosynthesis to fuel nitrogen fixation may store those products in the form of the polymer polyhydroxybutyrate (PHB). This polymer accumulates naturally in bacteria of certain nodules, most notably those of soybean, but much less so in their close relatives, such as *Phaseolus vulgaris* (French bean, navy bean) or species of *Vigna* (cowpea, green gram). Does this mean that soybean nodules and their bacteria are less well matched? Or does PHB have another function⁴? These are just two of the questions raised by the new results¹.

More broadly, other issues arise when we

Plant biology

Mutual sanctions

Janet Sprent

The bacterium-filled nodules found on legumes represent a mutually beneficial arrangement. But it is evidently one with sophisticated checks and balances to ensure a fair deal for both partners in the marriage.

Some soil bacteria live in apparent harmony with plant cells in a mutually beneficial arrangement. The bacteria can reduce nitrogen gas, 'fixing' it into forms that the plants can use; in return, the plant cells provide the bacteria with products of photosynthesis. On page 722 of this issue¹, Lodwig and co-workers describe an exchange-control system that enables the two partners to share their resources without either one becoming dominant.

The enzyme complex involved in nitrogen fixation is nitrogenase, which is ancient and widespread among bacteria. Nitrogenase can use a variety of substrates, but its main role in today's world is the production of

ammonia (NH₃) from nitrogen gas (N₂).

Whereas free-living nitrogen-fixing bacteria use ammonia for their own growth, those living in symbiosis with other organisms, such as in the nodules on roots of pea and bean plants, normally hand it over to their host in the form of ammonium ions, in exchange for products of photosynthesis that are used to provide the energy for nitrogen reduction². Why should these bacteria behave so altruistically, when by so doing they lose their own source of amino acids?

Lodwig *et al.* propose an answer. Using plants of the garden pea, they induced the formation of root nodules containing either wild-type or mutant nitrogen-fixing



Figure 1 Nodule variety. a–c, The nodules formed by nitrogen-fixing bacteria come in various forms, ranging from spherical, to branched and coraloid. The plants involved are, a, *Centrosema angustifolium*, a tropical forage legume; b, *Chadsia grevei*, a shrub from Madagascar; and c, *Enterolobium cyclocarpum*, a Brazilian tree. d, Nodules usually form on roots but on some species, such as *Aeschynomene* sp. from Senegal, shown here, they occur on stems.

look at the full landscape of nodulation processes. Our detailed knowledge of nodulation comes from just a few species of the more highly evolved legumes, mainly from temperate or sub-tropical regions. But legumes are the third largest family of flowering plants, and nodulation has arisen in them on several separate occasions during evolution; many woody species still lack this ability⁵. There are thus wide variations in the specificity and strength of the association with rhizobia, especially in the tropics (Fig. 2).

Many interactions lack the close co-evolution of host and bacteria that leads to the highly effective recognition and developmental processes evident in pea and most temperate species. A single host species may be nodulated by several different genera and species of bacteria⁵, with bacteria inside the nodules varying from the essentially parasitic

to the highly effective in delivering ammonia⁶. Similarly, a single strain of bacterium may nodulate many genera and species of legume. Bacteria that induce nodules are now known to be far more heterogeneous than once thought, with many having close relationships with plant or animal pathogens — even to the extent of being members of the same genus, as occurs, for example, with species of *Burkholderia* and *Ralstonia*. Symbiotic and pathogenic relatives may have similar ways of avoiding their host's defence responses⁷. Genetic exchange between bacteria in soil may lead to some species losing the genes determining symbiosis and nitrogen fixation, and others gaining these genes⁸. We can expect many 'new' nodulating bacteria to be found in the future.

When coupled with the impressive range of techniques for studying whole genomes

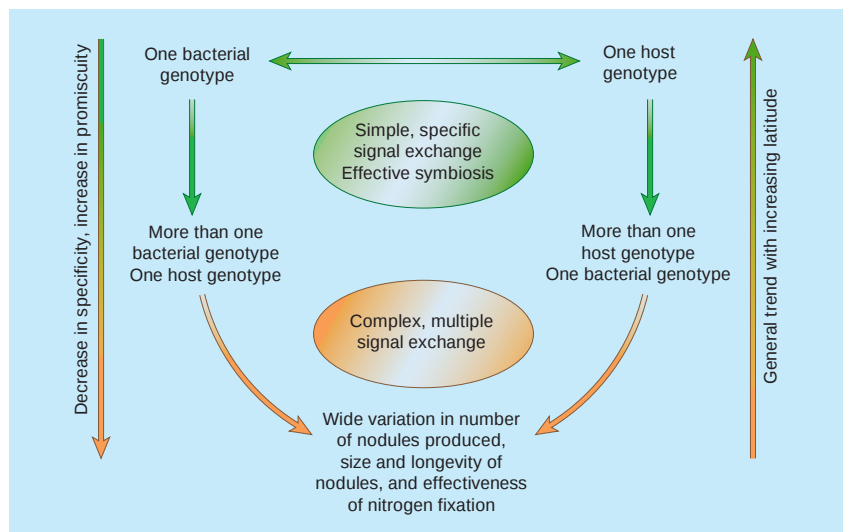


Figure 2 No fixed relationship. Interactions between soil bacteria and host legumes vary widely in their specificity, and in the strength of the association and its results. The most widely studied interactions are the highly specific ones found in advanced legumes of a particular subfamily (the Papilionoideae)⁵. But the less specific associations found in many legumes from all subfamilies may be more common. There is an overall trend in specificity and likelihood of nodulation with latitude (and, to some extent, altitude): the higher the latitude, the more specific the relationship between the host plant and the bacterium.



100 YEARS AGO

The Corporation of the City of London is rightly taking part in the crusade against tuberculosis. It has for many years instituted legal proceedings against farmers, butchers and meat-salesmen for sending tuberculous meat into the City markets, or for exposing the same for sale. Since it would appear that in some cases such offences may have been due to ignorance, the Public Health Department has issued a circular describing the indications of tuberculosis in the carcase, and the symptoms of the disease in the living animal.

ALSO...

Reuter reports that an eruption of the volcano Del Tierra Firme (Columbia), near Galera de Zamba, occurred on March 22 by which the village of Tiojo was destroyed. Brightly illuminated clouds, giving rise to the appearance of flames, were seen above the volcano on the night of March 24 by ships passing sixty miles off the coast. From *Nature* 16 April 1903.

50 YEARS AGO

At all periods, mankind has danced to get rid of surplus nervous emotion — to obtain release. During the First World War a United States hospital unit took over a British general hospital soon after the Germans had launched mustard-gas attacks. The sights and sounds were particularly distressing and the nurses, new to war conditions, in many cases became hysterical, though doing their duties magnificently. When the matron organized dances, the nervous tension was released and the troubles ceased... Even in prehistoric times it would seem that dancing had a place in the various cults. Both in ancient Egypt and in Greece dancers are shown in the pictures of religious festivals; again, we read in the Old Testament how David danced before the Lord... Naturally, then, when Europe became Christian, dancing was absorbed into the new cultus, though the Church naturally looked on it with disfavour, and from time to time attempts to exclude it were made. Nevertheless, it was not only the populace who frequently expressed their religious emotions by dancing; the clergy and choir, too, sometimes danced during the services. The Easter dances before the high altar in the cathedral at Seville are well known and still take place, and many less famous though equally ancient ones still happen in churches or churchyards at certain times of the year. From *Nature* 18 April 1953.

of legumes and other organisms⁹, and the detailed literature on the various steps in nodulation², highly targeted work such as that of Lodwig *et al.* will deepen our understanding of how nitrogen-fixing symbioses function. If this is extended to other legumes and other nodulating bacteria, exciting prospects are raised for answering questions ranging from why some legumes cannot nodulate to what distinguishes a pathogen from a symbiont. Above all, perhaps, given their agricultural importance, a better understanding of tropical legumes will assist the management of nitrogen fixation in those areas of the world that need it most. ■

Janet Sprent is emeritus professor in the School of Life Sciences, University of Dundee, Dundee DD1 4HN, UK.
e-mail: jispren@aol.com

1. Lodwig, E. M. *et al.* *Nature* 422, 722–726 (2003).
2. Perret, X., Staehelin, C. & Broughton, W. J. *Microbiol. Mol. Biol. Rev.* 64, 180–201 (2000).
3. Downie, J. A. & Parniske, M. *Nature* 420, 369–370 (2002).
4. Lodwig, E. & Poole, P. *CRC Rev. Plant Sci.* 22, 37–78 (2003).
5. Sprent, J. I. *Nodulation in Legumes* (Royal Botanic Gardens, Kew, 2001).
6. Burdon, J. J., Gibson, A. H., Searle, S. D., Woods, M. J. & Brockwell, J. J. *J. Appl. Ecol.* 36, 398–408 (1999).
7. Roop, R. M. II *et al.* *Vet. Microbiol.* 90, 349–363 (2002).
8. Van Elsas, J. D., Turner, S. & Bailey, M. J. *New Phytol.* 157, 525–537 (2003).
9. Trevaskis, B. *et al.* *Comp. Funct. Genom.* 3, 151–157 (2002).

extensive regions of continents, known as 'shields', that are older than 1 billion years. In contrast, the oldest oceanic material is only about 200 million years old, because of the cycle in which oceanic crust is created at mid-ocean ridges and subsequently destroyed as material is returned to depth in subduction zones (particularly around the Pacific). The preservation of old material as the continents move across the Earth's surface due to the relative motions of the tectonic plates is related to what lies beneath: samples brought to the surface through various eruptive processes indicate that there is a significant difference between the continental and oceanic environments.

Based on information from heat flow, geochemistry and the relative delay times of seismic waves in different settings, Jordan² proposed the 'tectosphere' model, in which a zone moves with the motion of the plate lying beneath the old continental shields and would be expected to be about 400 km thick. More recent assessments of heat-flow data and geochemistry favour a zone no thicker than 250 km. A thickness of 200–250 km is also consistent with investigations of how the Earth has responded to the removal of the load caused by glaciers, and with regional seismological studies using seismic surface waves. But many seismological models of three-dimensional structure based on global observations would favour a zone extending

Earth science

Roots of the matter

B. L. N. Kennett

How far down does the ancient continental material that constitutes Earth's 'tectosphere' extend? Fresh interpretation of the behaviour of seismic waves helps in reconciling previous estimates.

Over the past three decades there has been vigorous debate over how thick the continents can be — that is, the depth to which the rigid crust and upper mantle reach before meeting convecting mantle that can flow and drive tectonic

motion. On page 707 of this issue¹, Gung and colleagues add new seismological interpretations that go some way to explaining the differing views.

The oldest continental rocks are more than 3.8 billion years old and there are

Materials science

Mind the graphite gap

Graphite has diverse applications, ranging from pencils to electronic devices and nuclear reactors. Defects — displacements of carbon atoms — can occur in its structure, which may be beneficial in electronics, but could be dangerous in old-style, air-cooled nuclear reactors such as that pictured. There, the defects store energy and can lead to fire. Quantum-mechanical computer simulations by Rob Telling and colleagues (*Nature Materials* doi:10.1038/nmat876; 2003) now show that these defects may be structurally more complex than previously thought.

The structure of graphite itself is essentially simple. It consists of layered sheets known as graphene, each sheet being formed from a planar array of carbon atoms. Defects can form, for example, through the irradiation of graphite in nuclear reactors. This can induce a carbon atom to leave a sheet,

forming a 'vacancy defect'. Until now, it had been assumed that the remaining atoms in the sheet are unaffected, and retain a planar configuration.

But the simulations by Telling and colleagues show that the planar state would actually be unstable, and that the atoms surrounding the vacancy are more likely to be displaced out of the plane, very unlike the situation in a flawless sheet. The authors propose that if displaced atoms in two sheets were near to each other, a covalent bond could form between them, effectively bridging the gap.

A vacancy defect can also create another type of flaw, in which the removed carbon atom positions itself between two neighbouring sheets, forming an 'interstitial'. Again, covalent bonds could form. Because the interaction between sheets is usually quite weak — the



distance between them is 3.35 Å, some two-and-a-half times greater than the distance between atoms within a sheet — bonding between them has implications for the properties of the structure.

Although these simulations challenge current ideas about the nature of graphite defects, they are not inconsistent with some of the accepted evidence; they simply provide another explanation for the

results. Further research is needed to corroborate Telling and colleagues' theory, but, if validated, the new understanding may help to make the decommissioning of old nuclear reactors safer, and could pave the way to a whole new set of materials based on carbon nanotubes — effectively, rolled-up graphene sheets — which are already provoking great interest and a wealth of research. **Jane Morris**

BETTMANN/CORBIS

to 400 km, based on the depths at which seismic shear waves travel at elevated speeds — shear waves have their particle motion perpendicular to the direction of propagation. Such higher seismic wave speeds indicate the presence of a region possessing distinct properties, undoubtedly cooler than its surroundings, but also likely to be distinct in composition.

Gung and colleagues¹ help to reconcile these results. Their work is based on images of Earth structure derived from shear waves with different polarizations. The approach — seismic tomography — is akin to the medical tomography used to study the interior of the human body. In the seismic version, however, illumination of Earth's interior depends on the distribution of earthquakes generating seismic waves and of seismic stations producing high-quality data. Many different classes of wave-propagation phenomena are exploited to improve sampling deep into the Earth. The dominant energy from earthquakes is radiated as shear waves in which the deformation is transverse to their propagation path. Purely horizontally polarized shear waves have somewhat simpler characteristics than those with propagation in a vertical plane and so have been used in many studies. Gung *et al.* point out that the studies indicating a thick root to the continental shields make extensive use of data from the horizontally polarized waves.

When the authors segregate the information from vertically and horizontally polarized shear waves, they find that the resulting images show significant differences. The structure with fast wave speeds beneath most continental shields extends to at most 250 km for the vertically polarized waves, whereas the equivalent structures for horizontally polarized waves extend deeper, to nearly 400 km. Gung *et al.* suggest that this form of seismic anisotropy, with horizontally polarized shear waves travelling faster than those with vertical polarization, is due to the influence of mantle flow in the region between 250 and 400 km. Similar anisotropy has been found under the ocean basins^{3,4} in the depth range from 80 km to 250 km. The transition to anisotropy caused by mantle flow would then mark the seismic definition of the base of the 'tectosphere', with a depth of no more than 250 km beneath the ancient continental shields.

In most parts of the world it is difficult to provide any direct test of the results of Gung *et al.*¹ But the distribution of earthquakes around the shield of northern Australia, a region that I have studied, is such that higher-frequency seismic waves can be used to probe structure beneath the shield. Here, similar anisotropy is found, with faster travel of horizontally polarized waves⁵ through a zone below 220 km that produces significant dissipation of wave energy and is hence likely to have lowered viscosity, promoting flow. A

complicated structural transition to lowered seismic wave speeds beneath the ancient shield occurs at about 210 km and has itself been attributed to a different kind of anisotropy⁶. The patterns of seismic anisotropy are subtle and the strong effects associated with polarization are accompanied by variations with the direction of propagation⁷.

Overall, Gung *et al.* have provided a satisfying way forward in tackling a long-standing puzzle in Earth science. But on the evidence from northern Australia — and as might be expected — the picture will not be straightforward. The seismically defined base of the tectosphere is likely to be quite

complex, depending on variations in the local tectonic environment. ■

B. L. N. Kennett is in the Research School of Earth Sciences, The Australian National University, Canberra, ACT 0200, Australia.

e-mail: brian@rse.anu.edu.au

1. Gung, Y., Romanowicz, B. & Panning, M. *Nature* 422, 707–711 (2003).
2. Jordan, T. H. *Rev. Geophys. Space Phys.* 13, 1–12 (1975).
3. Ekström, G. & Dziewonski, A. M. *Nature* 394, 168–172 (1998).
4. Montagner, J.-P. *Rev. Geophys.* 32, 135–137 (1994).
5. Tong, C., Gudmundsson, O. & Kennett, B. L. N. *J. Geophys. Res.* 99, 15783–15797 (1994).
6. Leven, J., Jackson, I. J. & Ringwood, A. E. *Nature* 289, 234–236 (1981).
7. Debayle, E. & Kennett, B. L. N. *Earth Planet. Sci. Lett.* 184, 339–351 (2000).

Immunology

Oxygen and the inflammatory cell

Carl Nathan

The discovery that a single protein allows certain immune cells both to respond to low oxygen levels and to induce inflammation may provide a new target for drugs to treat diseases characterized by excessive inflammation.

Infected tissues, wounds, rheumatic joints, and parts of tumours that have outgrown their blood supply would seem to have little in common. Yet such sites share two features: they have lower concentrations of oxygen than healthy tissues (they are 'hypoxic'), and they are infiltrated by leukocytes, major cell types of the innate immune system. Writing in *Cell*, Cramer and colleagues¹ propose that a single protein mediates both the response of these cells to hypoxia and their ability to participate in inflammation — a coordinated immune response to tissue injuries such as those mentioned above². The protein, hypoxia-inducible factor-1 α (HIF-1 α), regulates the expression of at least 30 genes when oxygen levels are low³. Cramer *et al.* show that HIF-1 α also controls several key aspects of inflammation: the redness and swelling of injured tissues, and the ability of leukocytes to enter these sites. It is striking that a single molecule should emerge as a master regulator in two such diverse and significant settings as hypoxia and inflammation.

Leukocytes come in various guises, two major subtypes being neutrophils and macrophages. In one of the earliest responses to injury, neutrophils in an affected tissue's venules — small blood vessels that receive input from capillaries — stop flowing with the blood. Instead, they stick to the inside of the venules and form clumps that can grow to occlude the vessel, reducing blood flow. In addition, some neutrophils cross the vessel wall, migrating into the tissue. The reduced blood flow and increased cell numbers lead to local depletion of oxygen. Then, macrophage precursor cells arrive.

Hypoxia causes macrophages, neutro-

phils and other cells to activate a variety of their genes, including that encoding vascular endothelial growth factor (VEGF), which makes blood vessels leaky⁴. This gene activation is achieved by means of HIF-1 α , which binds to the promoters (regulatory elements) of target genes. To do so, HIF-1 α must first bind its partner, HIF-1 β , and this interaction is controlled by oxygen levels. When oxygen is abundant, there is little HIF-1 α — it is destroyed under the direction of the von Hippel-Lindau (VHL) protein — and what there is can't bind HIF-1 β . At low oxygen levels these twin restraints are lifted⁵ (Fig. 1, overleaf).

Given that HIF-1 α is active in hypoxic leukocytes, Cramer *et al.*¹ wanted to find out whether it is needed in inflammation. To study the function of a protein, researchers often inactivate its gene in experimental organisms such as mice, to see what happens. But disrupting the HIF-1 α gene kills mice when they are embryos, precluding any subsequent study of inflammation. So Cramer *et al.* used a special technique in mice to delete the protein only in leukocytes. The approach involved using the promoter of the lysozyme gene, which is largely active only in leukocytes, to switch on an enzyme that inactivates the HIF-1 α gene.

The researchers found that isolated macrophages and neutrophils from the mutant mice had reduced levels of VEGF and phosphoglycerate kinase compared with wild-type cells. The latter enzyme is needed to generate ATP, the main cellular energy store, from glucose by glycolysis; this oxygen-independent process is the predominant means by which leukocytes generate energy. As might be expected,

the leukocytes' ATP levels were also low.

In vitro, isolated macrophages showed defective migration and a reduced ability to kill bacteria. Moreover, mice with HIF-1 α -deficient leukocytes developed almost no redness or swelling in response to chemical irritants or arthritis-causing protein complexes, and fewer leukocytes infiltrated inflammatory sites (Fig. 1). By contrast, when the authors generated mice whose leukocytes lacked VHL, the inflammatory response was greatly increased, presumably because, without VHL, HIF-1 α was no longer destroyed. Cramer *et al.* conclude that HIF-1 α is needed for leukocytes to generate ATP in the low-oxygen conditions of injured tissues, and hence for these cells to function. These remarkable findings suggest that HIF-1 α might make a good target for drugs aimed at reducing the excessive inflammation associated with many diseases.

Does the decrease in VEGF expression in HIF-1 α -deficient leukocytes contribute to the reduced inflammatory response? Cramer *et al.*'s findings suggest that it might be partly responsible: they generated mice lacking VEGF, and found that the animals developed less swelling in injured sites (although they still showed extensive leukocyte infiltration). But the reduced swelling and redness associated with HIF-1 α mutant leukocytes might instead, or additionally, indicate a deficiency in the activity of mast cells. Mast cells are components of the innate immune system that act as sentinels stationed around blood vessels and beneath mucosal surfaces. They orchestrate the early phases of inflammation, including redness, swelling and leukocyte recruitment², as the arthritis model used by Cramer and colleagues shows⁶. Like leukocytes, mast cells can express lysozyme, and so may also have been lacking HIF-1 α in the mutant mice. If so, these cells might also have had decreased ATP levels and been sluggish in triggering inflammation. This again hints at the potential benefits of manipulating HIF-1 α to treat inflammation.

That idea might seem to have a drawback: HIF-1 α -deficient macrophages are less efficient at killing bacteria *in vitro*¹, suggesting that dampening inflammation might come with the price tag of a diminished ability to fight infection. But the situation could be different *in vivo*, and here it is interesting to look at nitric oxide (NO). Inflammation activates a gene that encodes the high-output form of NO synthase, an enzyme that makes NO; hypoxia, via HIF-1 α , contributes to this activation⁷. NO then promotes swelling and other aspects of inflammation, and helps macrophages kill microbes. It also mediates the avoidance and dormancy responses of fruitfly larvae to hypoxia⁸, and controls the number of mitochondria — the cell's energy-producing organelles — in mouse tissues⁹.

Moreover, in hypoxic conditions, NO inhibits cytochrome oxidase¹⁰, an enzyme

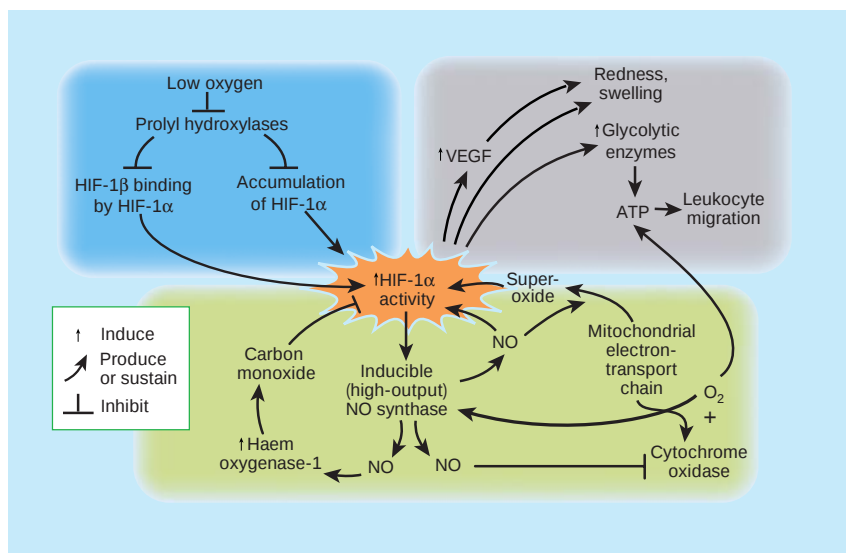


Figure 1 Factors that may control, or be controlled by, hypoxia-inducible factor-1 α (HIF-1 α) in inflammatory sites with low oxygen levels. The blue shaded area shows that hypoxia inhibits prolyl hydroxylase enzymes, which would otherwise restrict HIF-1 α activity. The grey area shows the findings of Cramer *et al.*¹: HIF-1 α is needed for several aspects of inflammation, namely the redness and swelling of injured tissues and, via glycolytic enzymes, leukocyte migration into injured areas. As shown previously, HIF-1 α also induces the production of vascular endothelial growth factor (VEGF). The green area shows additional considerations, taking into account that HIF-1 α increases the production of nitric oxide (NO). Specifically, NO can induce haem oxygenase-1, which produces carbon monoxide; this in turn inhibits HIF-1 α . But NO also increases HIF-1 α activity, both directly, and indirectly via the production of superoxide (through its effect on mitochondrial electron transport). And NO inhibits cytochrome oxidase, leading to reduced oxygen-dependent ATP synthesis, which leaves some oxygen available for further production of NO.

that is essential for oxygen-dependent ATP generation by mitochondria; other products of NO synthase impair ATP formation by glycolysis. So, in wild-type leukocytes under hypoxic conditions *in vivo*, ATP levels might fall farther than would be predicted from the decreased oxygen supply alone. Conversely, HIF-1 α -deficient leukocytes might express less NO synthase than wild-type cells, so their ATP generation could be less impaired. That suggests that ATP levels in HIF-1 α -deficient and normal leukocytes in hypoxic inflammatory sites might be more similar than they are in isolated cells cultured in air¹. If so, the cells might also be more similar in their ability to fight bacteria *in vivo* than *in vitro*.

Other interactions between NO and hypoxia deserve mention because they can affect the activation of HIF-1 α (Fig. 1). In hypoxic macrophages, if mitochondrial oxygen consumption by NO is blunted, the amount of intracellular oxygen available for other reactions¹¹, including further NO synthesis, will increase. When NO inhibits the mitochondrial transfer of electrons to oxygen, electrons leak out, eventually giving rise to superoxide. This species or its products can activate HIF-1 α ³. Furthermore, high-output NO synthase can stabilize HIF-1 α ¹². NO also induces haem oxygenase-1 (whose product, carbon monoxide, might inhibit HIF-1 α ¹³). On balance, low oxygen alone may not fully explain the activation of HIF-

1 α in inflammation. Instead, the combination of low oxygen and high NO may be important.

Researchers who seek to develop anti-inflammatory treatments based on this new understanding of HIF-1 α might be discouraged by the difficulty of inhibiting a protein whose function is not enzymatic, and whose known enzymatic regulators (the prolyl hydroxylases) inhibit rather than activate it. But let us hope that investigators will continue to probe the relationship between hypoxia and inflammation that has been highlighted in such a surprising way by Cramer and colleagues. ■

Carl Nathan is in the Department of Microbiology and Immunology and Graduate Programs in Immunology and Molecular Biology, Weill Medical College, Cornell University, 1300 York Avenue, New York, New York 10021, USA.

e-mail: cnathan@med.cornell.edu

1. Cramer, T. *et al.* *Cell* 112, 645–657 (2003).
2. Nathan, C. *Nature* 420, 846–852 (2002).
3. Semenza, G. L. *Trends Mol. Med.* 7, 345–350 (2001).
4. Senger, D. R. *et al.* *Science* 219, 983–985 (1983).
5. Ivan, M. *et al.* *Science* 292, 464–468 (2001).
6. Lee, D. M. *et al.* *Science* 297, 1689–1692 (2002).
7. Melillo, G. *et al.* *J. Exp. Med.* 182, 1683–1693 (1995).
8. Teodoro, R. O. & O'Farrell, P. H. *EMBO J.* 22, 580–587 (2003).
9. Nisoli, E. *et al.* *Science* 299, 896–899 (2003).
10. Moncada, S. & Eruslimsky, J. D. *Nature Rev. Mol. Cell Biol.* 3, 214–220 (2002).
11. Trimmer, B. A. *et al.* *Science* 292, 2486–2488 (2001).
12. Sumbayev, V. V. *et al.* *FEBS Lett.* 535, 106–112 (2003).
13. Huang, L. E., Willmore, W. G., Gu, J., Goldberg, M. A. & Bunn, H. F. *J. Biol. Chem.* 274, 9038–9044 (1999).

Developmental biology

The trouble with the sex body

Dev. Cell 4, 497–508 (2003)

The route by which male mammals make sperm has several stages, one of which requires the formation of a 'sex body'. This structure has quite a history in biology: it was first described in 1898, and a quarter-century later was recognized as being the closely associated X and Y chromosomes.

The genes on both chromosomes must be inactivated at this point as part of successful sperm formation. Oscar Fernandez-Capetillo and colleagues have been looking into what can go wrong in this process. Male (but not female) mice that lack a gene encoding a protein called H2AX are infertile. Fernandez-Capetillo *et al.* now find that the specific effects of lack of H2AX are failure to form the sex body and to inactivate the sex chromosomes. These consequences are likely to be associated with the protein's function in DNA packaging.

One issue to be resolved is the importance of phosphorylation of H2AX in events. Another is that, in other circumstances, the phosphorylated protein has been implicated in rejoining DNA sequences. Such DNA repair is an essential part of the general process of meiosis, by which both sperm and eggs are formed, so it seems odd that phosphorylated H2AX does not have that function here — or so it would appear, given that female mice that lack it are fertile.

Tim Lincoln

Microelectronics

Clean-up for copper chips

J. Am. Chem. Soc. doi:10.1021/ja034091m (2003)

Copper is increasingly used for the wiring in microchips, but patterning and smoothing it is a messy business. A technique described by Carol A. Bessel and colleagues promises to do away with the toxic wastes associated with liquid etching and polishing procedures that involve water or organic solvents. It uses as the solvent pressurized, supercritical carbon dioxide, which reverts to a gas when the pressure is eased. So this is essentially a 'dry' process with no liquid waste.

At present, etching of copper wiring in microcircuits involves a mixture of chemical removal and mechanical grinding with abrasives. This 'chemical mechanical planarization' technique requires a slurry of various ingredients that cannot be easily recycled. When water is the solvent, it can also damage other chip components.

Bessel *et al.* use chemical oxidants and chelating agents dissolved in supercritical CO₂ to etch copper metal at 40 °C and

a pressure of 214 bar. Up to 200,000 atomic layers of copper were removed from disks of the metal in 20 hours. But it remains to be seen whether the resulting surface can be made as smooth as chip manufacture demands.

Philip Ball

Plant biology

Ice on top

Genes Dev. doi:10.1101/gad.1077503 (2003)

Many plants can tolerate a sharp frost as long as they have the chance to adjust to low, non-freezing temperatures first. A chilly spell activates genes that are crucial for the adaptation period, as well as those needed to survive frosts. Viswanathan Chinnusamy *et al.* now show that, in the thale cress *Arabidopsis*, the transcription factor ICE1 lies at the top of a molecular cascade that turns on many of these cold-responsive genes.

Besides a good dose of antifreeze protein, a plant's cold-survival kit contains enzymes involved in metabolism, as well as so-called chaperone molecules and proteins that help it tolerate the dehydration caused by freezing. A family of transcription factors known as CBFs drives the collective production of such proteins.

The CBFs themselves are switched on exclusively at cooler temperatures — so how is this process controlled? Chinnusamy *et al.* have identified the 'inducer of CBF expression-1' (ICE1), which is a transcription factor present in most plant tissues at normal temperatures. The researchers speculate that exposure to cold causes ICE1 (or an associated protein) to become modified, allowing it to activate CBF transcription.

Various other cold-sensitive *Arabidopsis* mutants have already been isolated. So this may be just the tip of the iceberg in the regulation of cold-responsive genes.

Marie-Thérèse Heemels

Spongiform encephalopathies

Propagating warped prions

Cell 113, 49–60 (2003)

The long-standing theory that prion proteins couple up to breed degenerative brain diseases — such as bovine spongiform encephalopathy (BSE) in cows and scrapie in sheep — has gained some support. Philipp Meier *et al.* have caught misshapen prions warping normal ones in mouse brains.

In animal encephalopathies and their human analogue, variant Creutzfeldt–Jakob disease (vCJD), diseased prions clump together in the brain, eventually destroying it. Misshapen prions were thought to latch onto and alter normal ones, but this liaison had not been seen in live animals — until now. Meier *et al.* genetically engineered mice

to carry a new, artificial version of prions with an antibody tag. Unlike the real thing, the labelled prions are easy to isolate from cells. The authors then infused the animals' brains with the wrongly folded prion proteins responsible for scrapie. They found tagged, artificial prions attached to scrapie prions.

Meier *et al.* speculate that the artificial prions could form the basis of a new therapy. Although they bind mutant prions, they resist being transformed themselves. Mice carrying the modified prion survived exposure to scrapie proteins for at least three months longer than normal mice.

Helen Pearson

Archaeology

Tall story

J. Archaeol. Sci. 30, 417–428 (2003)

Roman civil engineers knew a trick or two. In the third century AD they built a 1.67-km pipeline to carry water from a distant source across a broad valley to a tank that supplied the city of Aspendos, in modern-day Turkey. Calculations of the flow in this 'inverted-siphon' system by C. R. Orloff and Adonis Kassinos show how its design helped to keep the flow rate steady.

The pipeline is divided into three segments, separated by two tall towers on the valley floor where the water has to ascend and then descend again (see picture). These hurdles, the researchers show, would have helped to damp out oscillations that might have arisen each time the pipeline was opened to refill the tank. Such back-and-forth sloshing would have made the flow intermittent and might even have damaged the pipeline.

According to the Roman writer Vitruvius, successful siphon function depended on *colliquaria*. The meaning of this term is lost in time, but Orloff and Kassinos suggest that it might refer to the small holes that perforate some of the stone blocks in which the sections of pipe are bored. These holes would have allowed water and air to escape and thus helped to dissipate turbulent 'slugs' and pressure waves, restoring steady flow.

Philip Ball



Water works — one of the towers at Aspendos, with (right) remains of the aqueduct.

Sexual transmission of HIV in Africa

Other routes of infection are not the dominant contributor to the African epidemic.

The human immunodeficiency virus (HIV) epidemic in sub-Saharan Africa continues to grow at an alarming rate, with 3.5 million people infected last year alone¹. It has recently been proposed that heterosexual transmission cannot account for more than 35% of HIV incidence in this region, and that parenteral transmission, probably through unsterile medical practices, is the main route of infection². A key prediction of this hypothesis is that the epidemic history of HIV in sub-Saharan Africa is similar to that of other blood-borne pathogens. To test this, we compared prevalence estimates of HIV with those of hepatitis C virus (HCV), which has a far greater rate of parenteral than sexual transmission. We show that HIV and HCV have such different epidemic histories in sub-Saharan Africa that parenteral transmission is unlikely to be the main source of HIV infection.

We plotted comparable estimates of HIV and HCV prevalence in the general population for all countries in the world for which data are available (Fig. 1a). Countries are segregated according to whether their HCV/HIV prevalence ratio is greater or less than 1 (black line in Fig. 1a). This distinction is very conservative, given that HCV is about six times more infectious than HIV by parenteral transmission³. Most countries have a ratio of >1 , as expected from the higher global prevalence of HCV. However, countries with ratios of <1 are largely from sub-Saharan Africa (Fig. 1a, red circles) and, crucially, these contain the majority of HIV-infected individuals worldwide (as indicated by circle sizes in Fig. 1a; India is the exception). This observation is contrary to what would be expected if most cases of HIV infection were attributable to parenteral transmission. It cannot be argued that HCV was absent from sub-Saharan Africa at the onset of the HIV pandemic, because HCV genotypes 1, 2 and 4 have been endemic to west and central Africa for several centuries, and subtype 5a has been present in South Africa for at least 50 years^{4,5}.

To complement this cross-sectional prevalence analysis, we plotted the temporal change in HIV and HCV prevalence in South Africa, the country with the greatest number of HIV-infected individuals in Africa, and for which annual data are available (Fig. 1b). Crucially, HCV and HIV were at similarly low prevalences at the beginning of the 1990s. Since then, the increase in HIV prevalence among adults has been swift, rising from 0.7% in 1990 to 24.5% in 2000 (ref. 6), whereas estimates of HCV prevalence have remained relatively low over the same period (for data

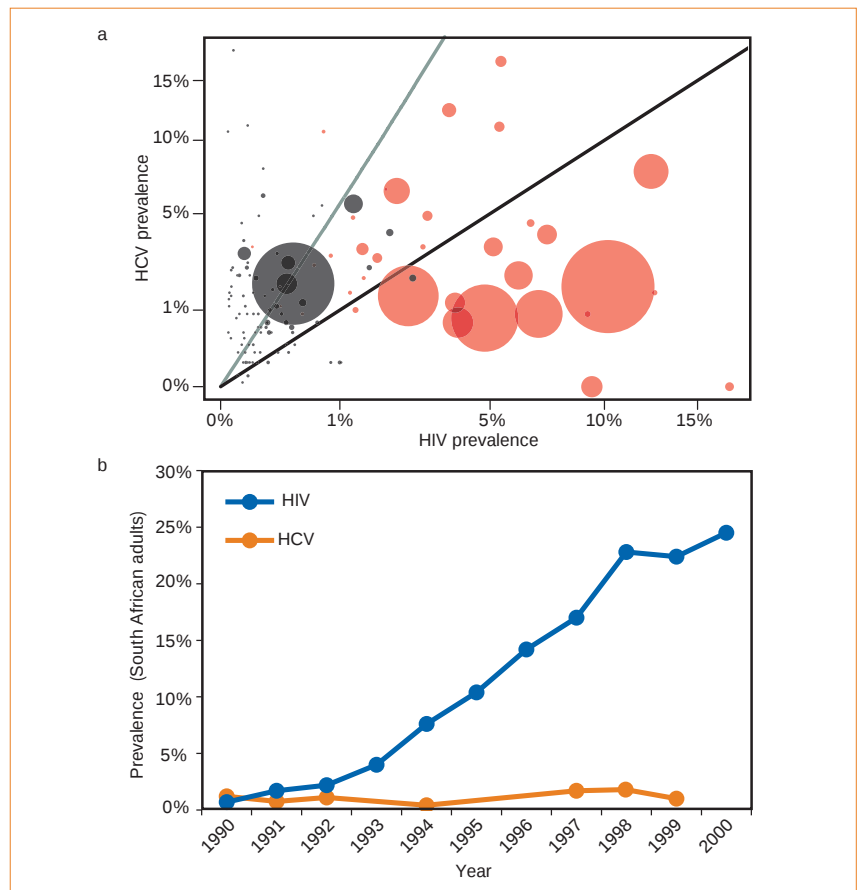


Figure 1 Comparison of epidemic histories of human immunodeficiency virus (HIV) and hepatitis C virus (HCV) in different countries, and in South Africa during the 1990s. **a**, HCV and HIV prevalence in the general population (including children) of every country for which data are available. The arcsine square-root transformation for proportions was used because the original distribution of points was strongly L-shaped. Sub-Saharan African countries are represented by red circles; other countries are represented by black circles; the radius of each circle is proportional to the number of HIV infected individuals in that country. Countries below the black line have a higher prevalence of HIV than HCV; the opposite is true for countries above the black line. The grey line represents an HCV:HIV prevalence ratio of 6, which indicates the relative parenteral transmissibility of the two pathogens³. All prevalence figures apart from five HCV values were obtained from World Health Organization sources. **b**, HIV and HCV prevalence in South African adults from 1990 to 2000 (estimates for HIV from ref. 6; estimates for HCV from sources listed in supplementary information).

sources, see supplementary information). The absence of a comparable increase in the prevalence of HCV, a virus that is more easily transmitted parenterally, suggests that current levels of HIV prevalence are not primarily the result of using unsterile medical equipment or contaminated blood products.

Although it is important to highlight the issue of pathogen transmission by unsafe injection, the different epidemic histories of HIV and HCV in sub-Saharan Africa indicate that parenteral transmission is not the dominant contributor to the African HIV epidemic. Consequently, the introduction of widespread drug treatment for HIV and cofactor infections, as well as continued efforts to encourage safe sex, are

urgently required to reduce the spread of HIV in Africa.

Polly R. Walker, Michael Worobey, Andrew Rambaut, Edward C. Holmes, Oliver G. Pybus
Department of Zoology, University of Oxford,
South Parks Road, Oxford OX1 3PS, UK
e-mail: edward.holmes@zoo.ox.ac.uk

- UNAIDS. Report on the Global HIV/AIDS Epidemic 2002 (2002).
- Gisselquist, D. & Potterat, J. J. *Int. J. STD AIDS* **14**, 162–173 (2003).
- Goldmann, D. A. *J. Allergy Clin. Immunol.* **110**, S21–S26 (2002).
- Smith, D. B. et al. *J. Gen. Virol.* **78**, 321–328 (1997).
- Pybus, O. G. et al. *Science* **292**, 2323–2325 (2001).
- South African Department of Health. *National HIV and Syphilis Sero-prevalence Survey of Women Attending Public Antenatal Clinics in South Africa 2000* (2001).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.

Space travel

Dual origins of light flashes seen in space

Light flashes are unusual visual phenomena that are observed in space and are caused by the interaction of energetic cosmic-ray particles with the human visual system. Using data gathered on board the Mir space station during the Sileye-2 experiment¹, we show here that there are two separate components of cosmic rays that cause these flashes: one due to heavy nuclei and one due to protons. This indicates that perception by an astronaut's visual apparatus could involve two complementary mechanisms.

Ever since light flashes in space were predicted² before the first space mission and subsequently reported by early Apollo astronauts, attempts to determine their cause have been made both in space and using ground-based particle accelerators³. As a result, several explanations have been proposed to account for the phenomenon (see ref. 4 and references therein).

The measured rate of occurrence of light flashes (LF) varies for different missions: on Mir⁵, it is much lower (Sileye-1, 0.18 ± 0.02 LF min⁻¹; Sileye-2, 0.13 ± 0.01 LF min⁻¹) than on Apollo⁶ (0.23 ± 0.1 LF min⁻¹), Skylab⁷ (1.3 ± 0.1 LF min⁻¹) and ASTP⁸ (0.46 ± 0.05 LF min⁻¹). This effect is probably due chiefly to a reduction in the speed of low-energy particles by Mir's hull material (aluminium more than 3 mm thick) and the equipment inside the craft.

As light flashes are caused by particles interacting with the human visual apparatus, their occurrence should be proportional to particle rate (the so-called 'latitude effect'). Particle rates outside the region of the South Atlantic Anomaly (SAA) depend on the geomagnetic cut-off, which is a function of the position-dependent geomagnetic field intensity and direction. The cut-off C represents the minimum rigidity for cosmic rays to reach Mir's orbit without being deflected outwards; it is lower at high geomagnetic latitudes (for Mir's orbit, the lowest cut-off is $C = 0.6$ gigavolts (GV)) and higher at the geomagnetic equator (where maximum cut-off is $C = 16$ GV). The lower the cut-off, the higher the rate of particles coming from outside the Earth's magnetosphere (we use the vertical cut-off at each location of Mir).

Light-flash and particle rates measured inside Mir were divided in 3-GV bins that separated the regions outside ($C \leq 18$ GV) and inside the SAA ($3 \leq C \leq 15$ GV, selected for Earth's magnetic field $B < 2.5 \times 10^{-5}$ tesla; Fig. 1). The light-flash rate, R_{LF} , is plotted against the particle rate, P , for all particles (almost exclusively protons) in Fig. 1a. In Fig. 1b, R_{LF} is plotted as a function of the rate of relativistic nuclei, $P_{\geq 6}$, with charge $Z \geq 6$, obtained by selecting particles

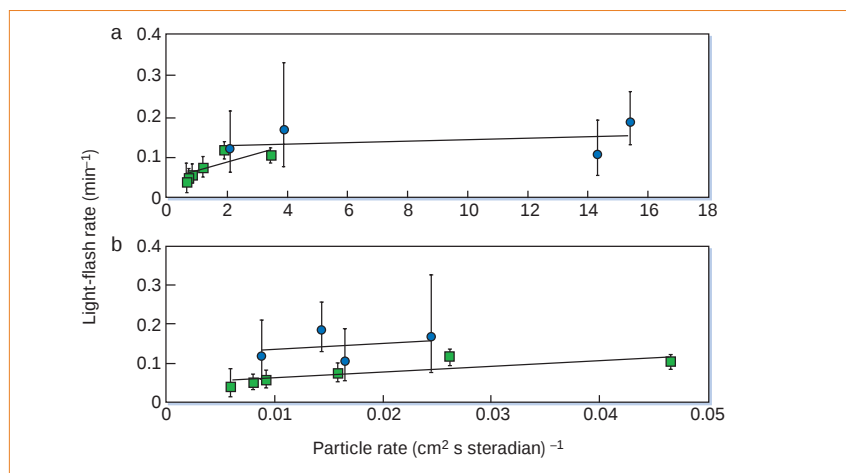


Figure 1 Rate of occurrence of light flashes on board the Mir space station as a function of particle rate for all particles and for relativistic nuclei inside (circles) and outside (squares) the South Atlantic Anomaly. **a**, Plot of light-flash rate against proton rate; **b**, light-flash rate against particle rate for particles with linear-energy transfer of > 20 keV μm^{-1} . Linear fits for each region are shown. Data are from the Sileye-2 experiment¹, in which astronauts wore light-excluding helmets integrated with cosmic-ray particle-flux detectors, enabling the frequency of flashes to be recorded as a function of background flux and orbit position.

with high linear-energy transfer (> 20 keV μm^{-1}) to guarantee the complete removal of the proton component.

From the all-particle plot (Fig. 1a), it is possible to see that R_{LF} is not linearly proportional to proton flux in all regions: in the SAA, it is roughly independent of proton rate (even though statistical errors preclude any further claim). For instance, at the centre of the SAA ($9 \leq C \leq 12$ GV), where particle rate is highest ($P = 15.4$ particles per cm^2 s steradian⁻¹, where steradian is a unit of spherical angle), the light-flash frequency is $R_{LF} = 0.18$ min⁻¹, which is only slightly higher than in other regions. This implies that the light flashes cannot be caused only by protons in all regions.

Figure 1b shows that frequencies of light flashes within the SAA are consistently higher than those for equivalent particle rates outside it, indicating that a further component must contribute to light flashes inside the SAA. Summing all observations in the SAA to reduce statistical errors, we obtain $R_{LF} = 0.15 \pm 0.03$ min⁻¹, compared with a rate outside it (in the same cut-off interval of the SAA) of $R_{LF} = 0.06 \pm 0.01$ min⁻¹. If heavy nuclei were the only cause of light flashes, these two values should overlap. The existence of two complementary mechanisms for light-flash perception could explain these observations: direct interaction of heavy nuclei with the retina, causing ionization or excitation, and proton-induced nuclear interactions in the eye (with a lower interaction probability) producing knock-on particles.

This explanation is also consistent with previous proposals^{3,6-9} and the fact that Mir

data inside the SAA are halfway between Skylab's (112 light flashes in 12 min in the SAA) and those of Apollo-Soyuz (no light flashes in the SAA). This is due to the average orbital height and shielding of Mir with respect to the other two space stations. Our results provide further insight into the overall effects of the space environment on humans.

M. Casolino et al.*

Department of Physics, University of Tor Vergata, 00133 Rome, Italy
e-mail: casolino@roma2.infn.it

*V. Bidoli¹, A. Morselli¹, L. Narici¹, M. P. De Pascale¹, P. Piccozza¹, E. Reali¹, R. Sparvoli¹, G. Mazzenga², M. Ricci², P. Spillantini³, M. Boezio⁴, V. Bonvicini⁴, A. Vacchi⁴, N. Zampa⁴, G. Castellini⁵, W. G. Sannita⁶, P. Carlson⁷, A. Galper⁸, M. Korotkov⁹, A. Popov⁹, N. Vavilov⁹, S. Avdeev⁹, C. Fuglesang¹⁰

¹Department of Physics, University of Tor Vergata, 00133 Rome, Italy; ²LNF-INFN, Frascati, Rome, Italy; ³Department of Physics, University of Florence, Italy; ⁴Department of Physics, University of Trieste, Italy; ⁵TROE of CNR, Florence, Italy; ⁶Neurophysiology-DISMR, University of Genova, Italy, and Department of Psychiatry, State University of New York, Stony Brook, New York, USA;

⁷Royal Institute of Technology, Stockholm, Sweden; ⁸Moscow State Engineering Physics Institute, Moscow, Russia; ⁹Russian Space Corporation Energia Korolev, Moscow, Russia; ¹⁰European Astronaut Centre, Cologne, Germany.

1. Bidoli, V. et al. *J. Phys. G* **27**, 2051–2064 (2001).
2. Tobias, C. A. *J. Aviat. Med.* **23**, 345–372 (1952).
3. McNulty, P. J., Pease, V. P. & Bond, V. P. *Science* **201**, 341–343 (1978).
4. Horneck, G. *Nucl. Track Radiat. Meas.* **20**, 185–205 (1992).
5. Avdeev, S. et al. *Acta Astronautica* **50**, 511–525 (2002).
6. Pinsky, L. S., Osborne, R. E., Bailey, J. V., Benson, R. E. & Thompson, L. F. *Science* **183**, 957–959 (1974).
7. Hoffman, R. A., Pinsky, L. A., Osborne, W. Z. & Bailey, J. Z. *NASA Report SP-377*, 127–130 (1977).
8. Budinger, T. F. et al. *NASA Report SP-412*, 193–209 (1977).
9. Charman, W. N., Dennis, J. A., Fazio, G. G. & Jelley, J. V. *Nature* **230**, 522–524 (1971).

Competing financial interests: declared none.

brief communications is intended to provide a forum for both brief, topical reports of general scientific interest and technical discussion of recently published material of particular interest to non-specialist readers. Priority will be given to contributions that have fewer than 500 words, 10 references and only one figure. Detailed guidelines are available on Nature's website (www.nature.com/nature) or on request from nature@nature.com

The extent and patterns of usage of Agent Orange and other herbicides in Vietnam

Jeanne Mager Stellman*, Steven D. Stellman†‡, Richard Christian§, Tracy Weber* & Carrie Tomasallo*

* Departments of Health Policy and Management and † Epidemiology, Mailman School of Public Health, Columbia University, 600 West 168th Street, New York, New York 10032, USA

‡ Institute for Cancer Prevention, One Dana Road, Valhalla, New York 10595, USA

§ 2102 Old Stage Road, Alexandria, Virginia 22308, USA

Herbicides including Agent Orange were sprayed by United States forces for military purposes during the Vietnam War (1961–1971) at a rate more than an order of magnitude greater than for similar domestic weed control. In 1974, the US National Academy of Sciences published estimates of the extent and distribution of herbicides sprayed. Here we present revised estimates, developed using more-complete data. The spray inventory is expanded by more than seven million litres, in particular with heavily dioxin-contaminated herbicides. Estimates for the amount of dioxin sprayed are almost doubled. Hamlet census data reveal that millions of Vietnamese were likely to have been sprayed upon directly. Our identification of specific military herbicide targets has led to a more coherent understanding of spraying. Common errors in earlier interpretations of the spray data are also discussed.

Between 1961 and 1971 herbicide mixtures, nicknamed by the coloured identification band painted on their 208-litre storage barrels, were used by United States and Republic of Vietnam forces to defoliate forests and mangroves, to clear perimeters of military installations and to destroy ‘unfriendly’ crops as a tactic for decreasing enemy food supplies¹. The best-known mixture was Agent Orange. About 65% of the herbicides contained 2,4,5-trichlorophenoxyacetic acid (2,4,5-T), which was contaminated with varying levels of 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD). Herbicide mixtures are listed in Table 1.

In 1970, the US Congress directed the US Department of Defense (DoD) to engage the National Academy of Sciences to conduct a comprehensive study (NAS-1974) of the ecological and physiological effects of defoliation in Vietnam². NAS-1974 relied on a chronological record, the HERBS file³, which contained flight path coordinates of Air Force spraying missions carried out between August 1965 and December 1971 and from 1968 on US Army helicopter spraying missions. In 1985, the DoD supplemented this file with the Services-HERBS file, derived from additional record searches. The HERBS file error rate was about 10%, attributable largely to transcription, data entry and pilot recording errors⁴. Under contract to the NAS, using data more complete than were then available, we undertook, in close collaboration with the US Armed Services Center for Research of Unit Records (CRUR), to correct both files (see Methods) and during this process discovered much additional archived data.

Military herbicide operations in Vietnam became a matter of scientific controversy from their inception^{5,6}. In April 1970, 2,4,5-T was banned from most US domestic uses, on the basis of evidence of its teratogenicity⁷. The Agent Orange Act of 1991 requested the Institute of Medicine (IOM) to assess the strength of the evidence for association between exposure to military herbicides and disease in veterans and the feasibility of conducting further epidemiological studies. The IOM recommended that the Department of Veterans Affairs develop historical reconstruction methods for characterizing exposure to herbicides in Vietnam^{8,9}. The present report is the result of that recommendation.

Background to military use of chemical defoliants

The DoD’s Advanced Research Project Agency’s (ARPA) Project Agile was instrumental in the US development of herbicides as a military weapon, an undertaking inspired by the successful British use of 2,4,5-T to destroy jungle-grown crops during the insurgency in Malaya. ARPA supported tests on combinations and concentrations of herbicides; calibration studies of the spray delivery system to achieve the desired 28 l ha⁻¹ (3 gallons/acre) rate; and experiments on optimal conditions to minimize spray drift¹⁰. ARPA also developed the Hamlet Evaluation System¹¹ which collected the political census data that we use here for estimating population exposures.

The first large-scale US military defoliation took place in Camp Drum, New York, in 1959, using Agent Purple (a 50:50 mixture of 2,4-D and 2,4,5-T) and a spray system which was the model for those used in Vietnam. Herbicide tests were run from August to December 1961 in the Republic of Vietnam (RVN), using dinoxol and tri-noxol^{12,13}. An insecticide test series was also undertaken. The first major herbicide shipment arrived in RVN in January 1962; defoliation targets were sprayed during September and October 1962 (Agent Purple); crop destruction targets were sprayed in November 1962 (Agent Blue)¹⁴. Systematic testing of herbicides and calibration of herbicide delivery systems continued for several years¹⁵.

A 1962 pact assigned ownership of the herbicides to RVN when they entered its territory. Vietnamese physically handled the herbicides during off-loading, transport, and transfer to storage tanks. RVN ownership complicated United States Air Force (USAF) logistics and record-keeping, and disposal when Agent Orange use was abandoned in mid-1970 (ref. 16). US policy emphasized that its forces were assisting the RVN in the herbicide programme. C-123 aircraft carried out the missions camouflaged and equipped with removable identification insignias. Crop destruction aircraft bore South Vietnamese markings and were accompanied by a Vietnamese crew member under a State Department/DoD concept known as *Farmgate*¹⁷. Flight crew wore civilian clothing.

The herbicide targets and US Air Force project folders

US Air Force (USAF) operations, codenamed Operation Ranch Hand, dispersed more than 95% of all herbicides used in Operation

Trail Dust, the overall herbicide programme. Other branches of the US armed services and RVN forces, generally using hand sprayers, spray trucks (Buffalo turbines), helicopters and boats, sprayed much smaller quantities of herbicide. Operation Ranch Hand was organized into projects that underwent a complex combined South Vietnamese and US approval system which could sometimes last as long as one year. Each project consisted of specific targets that were often amended or deleted during the approval process. Crop destruction also required White House approval until 1963, after which final approval was delegated to the US Ambassador to the RVN.

We reconstructed the project number to which each mission in the HERBS file belonged by concatenating two data fields. Aggregating missions by project number transforms the HERBS file from a chronological listing of criss-crossing flight paths into target-related groups of flights flown at different points in time (Figs 1 and 2). The importance of projects and targets has not been sufficiently appreciated. NAS-1974, and even the USAF itself¹⁸, inverted the hierarchy thus: "All missions within a target formed a project".

With US National Archives staff assistance we located a collection of USAF operational project folders for about 50% of the projects and nearly 60% of the volume of herbicide inventoried in the HERBS file¹⁹. Many folders contained 'after action reports' or other documentation on completed missions which had not been included in any HERBS file and permitted us to identify some 200 new missions that pre-date August 1965, the date of the earliest missions on the NAS-1974 HERBS file. The early years of the Vietnam War have incorrectly been regarded as of minor importance with regard to herbicide spraying. In total, about 1.9 million litres of Agent Purple were sprayed between 1962 and 1965, which is particularly significant because herbicides manufactured in the early 1960s were almost certainly more heavily TCDD-contaminated than those produced later²⁰. Further, pre-1965 spraying was limited to a relatively small area (Fig. 1c), which thus may be at particular risk for TCDD contamination. Recent data on TCDD residues in soil sampled near US Army Special Forces camps where Agent Purple was sprayed²¹, and results of soil assays at the testing grid in Eglin Air Force Base, Florida²², support this interpretation.

Revised spray inventory

We have re-estimated the volume and type of herbicides sprayed between 1961 and 1971 to have 7,131,907 more litres than the 'uncorrected' NAS-1974 inventory and 9,440,028 l more than NAS-1974's 'corrected' inventory, in which about 10% of all missions had been discarded because of obvious recording errors. Figure 3 shows the areas sprayed; Fig. 4 shows the yearly distribution of spray; Fig. 5 shows the numbers of sorties flown, 1961–1971. Table 2 gives estimates of the frequency with which land areas were repeatedly sprayed.

Contamination of 2,4,5-T with TCDD varied widely by production run, manufacturer, and the percentage 2,4,5-T in the formulation. In early 1966, Agent White, which did not contain 2,4,5-T and hence was not TCDD-contaminated, began to replace Agent Orange. Chemical market forces had led to a shortage of Agent Orange. From a tactical perspective, Agent White was less satisfactory than Agent Orange because several weeks were required for defoliation to begin. It was accepted by the DoD, however, because Agent Orange would apparently no longer be available in sufficient quantities²³. Agent Blue was the agent of choice for crop destruction by desiccation throughout the entire War, but more than four million litres of the other agents, primarily containing 2,4,5-T, were also used on crops.

Procurement records show that at least 464,164 l of Agent Pink and 31,026 l of Agent Green, with comparatively higher TCDD levels, were purchased, but we have been able to document little more than 50,000 l as having been sprayed in RVN and about 15,000 l that were used in tests. We have identified missions which dispersed about 1.9 million litres of Agent Purple, but available procurement data show purchase of only 548,100 l (ref. 24), indicating that the procurement records are incomplete.

Extent of dioxin contamination

Estimates of how much TCDD was deposited in Vietnam are based on the volume of 2,4,5-T-containing herbicide sprayed (which has been revised upward on the basis of the new inventory), and on TCDD contamination levels. Estimates of the mean contamination level used by the USAF¹⁸ and by NAS-1974 are probably too low.

Table 1 Use of military herbicides in Vietnam (1961–1971)^{13,15,29}

Name	Chemical constituents	Concentration active ingredient	Years used	Estimated quantities sprayed (litres)*
Agent Pink†	60%–40% n-Butyl: isobutyl ester of 2,4,5-T§	961–1,081 g l ⁻¹ acid equivalent‡	1961; 1965	50,312 sprayed; 413,852 additional on procurement records
Agent Green†	n-Butyl ester 2,4,5-T	(Should have same acid equivalent as Agent Pink)	(Unclear but within timeframe for Agent Pink)	31,026 shown on procurement records
Agent Purple†	50% n-Butyl ester 2,4,-D; 30% n-butyl ester 2,4,5-T; 20% isobutyl ester 2,4,5-T	1,033 g l ⁻¹ acid equivalent	1962–1965	1,892,773
Agent Orange†	50% n-Butyl ester 2,4,-D; 50% n-butyl ester 2,4,5-T	1,033 g l ⁻¹ acid equivalent	1965–1970	45,677,937 (may include Agent Orange II)
Agent Orange II †	50% n-Butyl ester 2,4,-D; 50% isooctyl ester 2,4,5T	910 g l ⁻¹ acid equivalent	After 1968 (?)	Unknown but at least 3,591,000 shipped
Agent White	Acid weight basis: 21.2% tri-isopropanolamine salts of 2,4-D and 5.7% picloram	By acid weight: 240.2 g l ⁻¹ 2,4,-D and 64.9 g l ⁻¹ picloram	1966–1971	20,556,525
Agent Blue (powder)	Cacodylic acid (dimethylarsinic acid) and sodium cacodylate	Acid: 65% active ingredient; salt: 70% active ingredient	1962–1964	25,650
Agent Blue (H ₂ O solution)	21% sodium cacodylate + cacodylic acid to yield at least 26% total acid equivalent by weight	Acid weight: 360.3 g l ⁻¹	1964–1971	4,715,731

Other chemicals used in testing programme but not in Vietnam operations include Modified Orange (4-amino-3,5,6-trichloropicolinic acid (picloram) added to Orange), Dalapon, Bromacil, Tandex, Monuron, Diuron and maleic hydrazide. Dinoxol (1890 l) (butoxyethanol esters of 2,4-dichlorophenoxyacetic acid (2,4-D) and 2,4,5-T), Trinoxol (1455 l) (40% ethanol ester of 2,4,5-T) and 378 l Conc D (30% ethyl ester of 2,4,-D in H₂O) were also used in tests during 1961.

*Nominal application rate: 4.78 kg ha⁻¹.

† Contaminated with varying levels of TCDD.

‡ Acid equivalent is the mass of pure acid that results from complete de-esterification or deamination of salts and esters. Total ester masses are approximately 20% greater.

§ 80–20% mixture when mixed with Agent Green. Agent Green was never sprayed alone but was immediately mixed with Agent Pink for spraying.

|| Proprietary product of Dow Chemical Company (Tordon 101).

¶ Ansil Chemical Co. product Phytar 560 was the only arsenical in use before July 1969.

After Agent Orange spraying ended in May 1970, the USAF was required to dispose of very large stockpiles of surplus herbicide that were ultimately incinerated aboard the M/T *Vulcanus* in 1977. (The US government had also assumed liability for large unblended inventories of 2,4,5-T and 2,4-D from contracted suppliers, either paying for storage or stockpiling the chemicals at Kelly Air Force Base, Texas, in December 1970 (ref. 25). We do not know the fate of the stockpiles.)

TCDD concentrations varied widely in the mixtures. The USAF was required to provide the US Environmental Protection Agency with an Environmental Impact Statement (EIS) prior to incineration²⁶. The EIS sheds light on TCDD levels in the over three million litres of herbicide stockpiled at the Naval Construction Battalion Center (NCBC), Gulfport, Mississippi, and the approximately 26,000 208-litre barrels in the Johnston Island stockpile.

TCDD concentrations ranged from 6.2 to 14.3 p.p.m., and averaged 13.25 p.p.m. in samples drawn for incineration-effluent modelling studies from 28 different NCBC barrels chosen by the USAF as representative of the seven manufacturers of the NCBC stockpile²⁶. However, in other samples drawn from the NCBC stockpile, the TCDD range was about 0.05 to 13.3 p.p.m. (weighted average 1.77 p.p.m.). (NAS-1974, however, calculated a range of <0.05 to 17.0 p.p.m., and an arithmetic mean of 2.99 p.p.m. from the same analytical data².) The large discrepancies between the effluent modelling study and the other samples analysed by the

USAF and NAS-1974 can be explained by information in the background analytical documentation in which runs analysed by NAS-1974 are consistent with the 'low dioxin' analytical series^{27,28}. The documentation also reports dioxin levels to be heterogeneous, even within production runs. USAF chemists had concluded that generalization from tested barrels to untested ones in the same production batch was not reliable. NAS-1974 appears not to have been provided information either on the heterogeneity of production runs or that they were generalizing from data consistent with a 'low dioxin' analytical series.

Two hundred random samples from the Johnston Island stockpile were also analysed for TCDD content. The Johnston Island stockpile is likely to have been primarily contracted for in 1967 or later and to consist almost exclusively of Agent Orange, because Agents Purple, Pink and Green were not manufactured after the use of Agent Orange began²⁹ and in mid-1966 Ranch Hand missions were curtailed because of a severe herbicide shortage¹⁷. NAS-1974 calculated a mean TCDD level of 1.91 p.p.m. $\pm$ 20% for the stockpile². USAF documentation that is widely viewed as authoritative¹⁸, however, disputes this mean and contends that the four highest values (17, 22, 33 and 47 p.p.m.) must have been Agent Purple, and not Agent Orange, because these values exceeded the mean reported for the NCBC inventory, citing a personal communication from a military officer who recalled that as many as 20 drums of Agent Purple may have been present in the stockpile and redrummed into

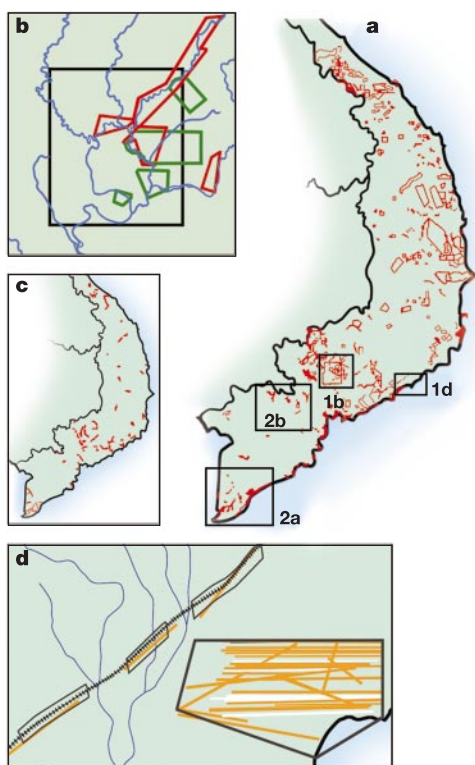


Figure 1 Herbicide projects, targets and spraying¹⁹. **a**, Spraying operations were directed at specific targets, 487 of which are shown. Labels in **a** refer to areas blown up in the corresponding parts of Figs 1 and 2. **b**, Some areas were targeted in multiple projects at different times: black box, 1964; red boxes, 1965; green boxes, 1966. **c**, We identified targets active before 1965; at least 4.7 million litres of herbicides were used to destroy 33,339 ha of crops and defoliate 101,300 ha of land. Dioxin contamination may be particularly relevant for this time period. **d**, Correspondence between HERBS file Agents Orange and White mission paths (orange and white lines) and target boxes, three of which enclose a railroad right-of-way. In Figs 1 and 2, waterways such as canals and rivers are shown in blue.

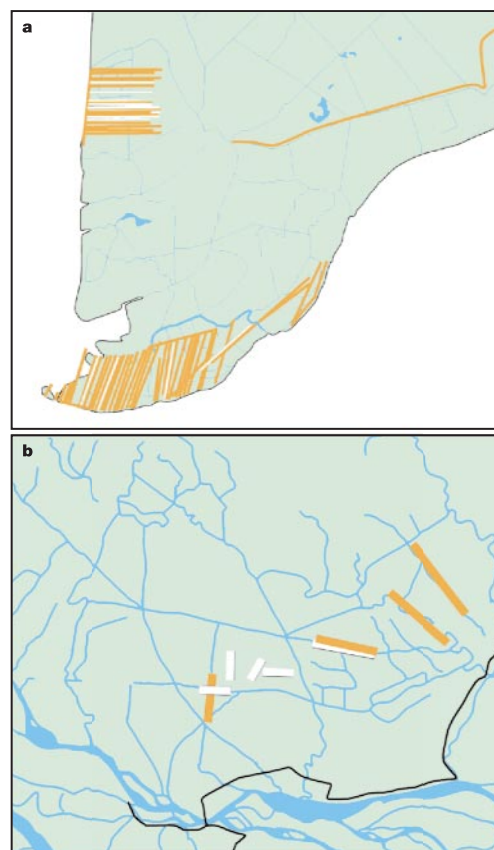


Figure 2 We could not locate target documentation for half the Ranch Hand missions on the HERBS file. **a**, Missions clearly directed at specific targets. Our current work is to identify probable targets through such mapping. Colours represent the flight paths of spray missions delivering Agents Orange and White **b**, HERBS file problem. Some purposes defined in the HERBS file documentation³, such as waterways and communication lines, are often labelled defoliation in the spray tapes, as shown. NAS-1974 and others have analysed spraying by HERBS file purpose but this is inaccurate.

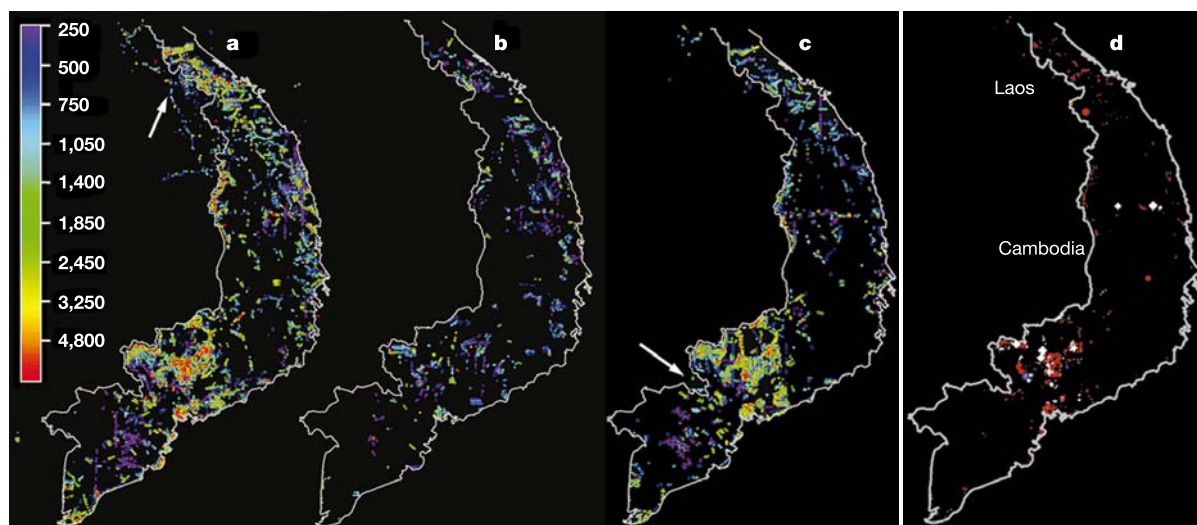


Figure 3 Volumes of herbicide sprayed. **a**, **b** and **c** represent known volumes of Agents Orange, White and Blue, respectively, sprayed by US military forces in RVN, 1961–1971. Volumes are calculated for individual grids spaced at 0.01 degrees ($\sim 1.2 \text{ km}^2$), that divide up Vietnam in a geographic information system developed by us⁴⁵. Colours in **a**, **b** and **c** correspond to volumes as shown in key. Arrows in **a** and **c** point to missions in Laos

and in the Parrot's Beak region of Cambodia, respectively. **d**, Grids sprayed with volumes greater than 4,800 l (about 10% of total), with marker size increasing in proportion to volume and colours corresponding to herbicide codenames. All herbicides containing 2,4,5-T are represented by orange markers.

Agent Orange containers. (The hypergeometric probability of selecting four of the 20 Agent Purple drums from the stockpile is 1.32×10^{-5} .) In fact, the range observed is completely consistent with the USAF's own analysis of the range and heterogeneity of TCDD levels^{27,28}. By 1988, Young, the senior author of the USAF documentation, dropped the word “may” and simply reported the four high values to have been Agent Purple³⁰. This latter reference has been relied upon as authoritative by the IOM⁸, and many others.

A 1971 NAS-1974 analysis of six core soil samples collected from the central calibration grid at Pran Buri, Thailand, over which all ARPA test flights had flown, found TCDD levels ranging from non-detectable ($<0.0012 \text{ p.p.m.}$) to 0.0233 p.p.m. and 2,4,5-T residue from non-detectable ($<0.02 \text{ p.p.m.}$) to 0.61 p.p.m. . NAS-1974 estimated the original herbicide to have contained <3 to 50 p.p.m. TCDD, consistent with the range observed in the Johnston Island stockpile². (The USAF documentation¹⁸ incorrectly asserts that NAS-1974 had erred in attributing the TCDD to Agent Orange rather than to Agents Purple and Pink. These misstated findings are used as further rationale for assuming the four high TCDD values to have been Agent Purple.)

Although Agent Purple is, indeed, likely to have been more highly contaminated with TCDD (an archived sample of Agent Purple at Eglin Air Force Base contained 45 p.p.m. TCDD¹⁸ and historical TCDD contamination data show early 1960s contamination levels to have been much higher²⁰), it is also likely that mean TCDD levels in Agent Orange were far higher than 3 p.p.m. for much of the herbicide used. An average value closer to 13 p.p.m. may be more realistic.

If 3 p.p.m. , the mean associated with the ‘low dioxin’ series is conservatively applied to the new inventory we have presented here, the estimate for TCDD present in the spray grows to 221 kg from NAS-1974 estimates of 106 – 163 kg . Applying 32.8 p.p.m. and 65.5 p.p.m. as the average TCDD in Agents Purple and Pink, we obtain an additional 165 kg , or 366 kg in total (which still does not take into account the herbicides sprayed by RVN forces, and possibly by US Army and Navy forces by trucks, boats, hand sprayers and helicopters, nor the more than $400,000 \text{ l}$ of Agent Pink shown in procurement records but not found in any recorded missions). If, indeed, dioxin contamination of Agent Orange could be fourfold or more higher, then this increased dioxin load grows

proportionally. It is also possible that some missions recorded as having dispersed Agent Orange did, in fact, spray the much more highly contaminated but unaccounted-for Agent Pink, but we know of no way to determine this. It is more likely that the unaccounted-for herbicides were used by Vietnamese troops, although about $50,000 \text{ l}$ of Agent Pink do appear on the 1965 inventory.

Estimates of population exposure to herbicides

A Hamlet Evaluation System (HES) in which US district advisors and Vietnamese district chiefs filled out monthly political survey and census forms was established in June 1967 and a gazetteer of place names and precise geographical locations was also created. The HES data provide a comprehensive rural census that permits us to estimate the numbers of hamlets and size of the population directly sprayed upon³¹. HES files were not made available to NAS-1974 early enough to permit analyses².

More than $20,585$ unique hamlets are represented in the corrected version of the database used here. Population data are not

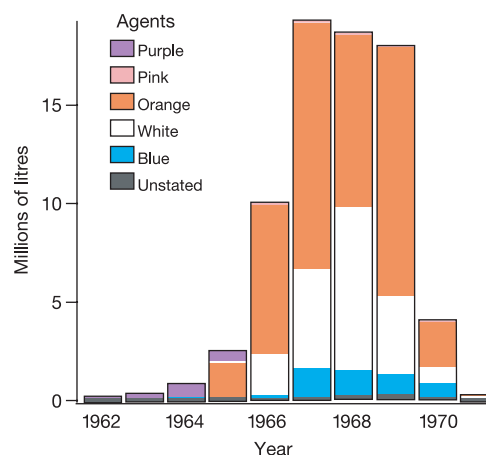


Figure 4 Litres of herbicides sprayed over 1962–1971. The data are taken from the corrected HERBS file and do not include Dinoxol and Trinoxol used in 1961.

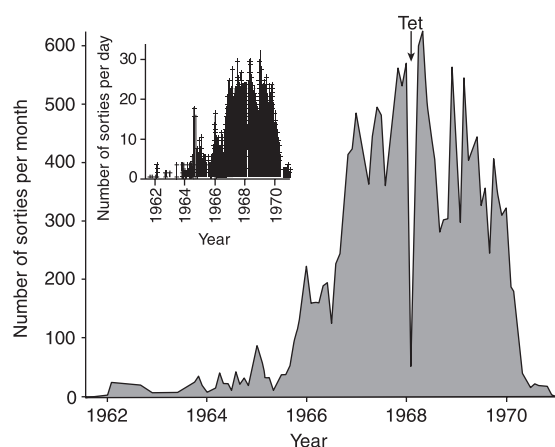


Figure 5 Time course of herbicide sorties. At least 19,905 sorties were run between 1961–1971 (1–34 daily, with a daily average of 10.7 sorties). These data disagree with USAF estimates¹⁸. The number of daily sorties mirrors the course of the Vietnam War itself: a slow build-up, maximum activity 1968–1969, then a slow but steady decline. The abrupt spraying drop-off at the end of January 1968 corresponds to the Tet Offensive wherein North Vietnamese forces carried out massive, coordinated attacks throughout RVN for nearly 2.5 months. Spray equipment was removed from Ranch Hand aircraft and crew and aircraft participated in airlift operations.

available for 18% of these hamlets and population data are not systematically reported each month for all years. Among the hamlets with some population data, 3,181 were sprayed directly and at least 2.1 million but perhaps as many as 4.8 million people would have been present during the spraying. Another 1,430 hamlets were also sprayed, but we cannot estimate the population involved. In all, at least 3,851 out of 5,958 known fixed-wing missions had flight paths directly over the hamlet coordinates given in the HES and gazetteer data and about 35% of the total herbicide sprayed was flown by these missions, although, in general, flight paths extended beyond hamlet borders.

Aborted missions and emergency dumps

Forty-two missions originally intended to spray 120,000 l of herbicide are known to have ended with emergency herbicide dumps where the chemical was jettisoned in about 30 s, compared to the usual 4 to 5 min. At least five herbicide-loaded aircraft crashed. Hundreds of other missions were aborted after take-off because of poor weather at the target site, heavy anti-aircraft fire, or mechanical problems. Such aircraft returned to base with herbicide load intact. It has been erroneously reported that aborted missions automatically dumped herbicide before landing³². Many flights were 'aborted' before take-off because of mechanical problems. One extensive but incomplete list of aborted missions is contained in a large-format uncorrected version of the HERBS file, known as the Map Book, which has served as the basis for much analysis of potential exposure, particularly by Vietnamese scientists³³. When the Map Book list is checked against the USAF Daily Air Activities Reports (DAARS), it is found to consist primarily of mechanical aborts, not herbicide dumps. Conversely, the list does not contain many actual dumps identified by CRUR and has other inaccuracies which have been corrected over the years by the DoD and now by us.

In 1971, NAS-1974 analysed five soil samples from an area in which about 3700 l of Agent Orange had been dumped in December 1968. No 2,4,5-T could be detected. TCDD was not analysed³⁴. Aborted missions may not represent the significant source of exposure that it has been represented to be by others, and studies which have relied on the uncorrected data in the Map Book will, of necessity, be inaccurate.

Table 2 Estimated area and frequency of spraying

Times sprayed	Hectares	
	All herbicides	2,4,5-T herbicides
1	368,556	343,426
2	369,844	332,249
3	361,862	275,770
4	341,037	236,232
5	272,709	153,192
6	216,724	119,127
7	153,391	75,062
8	138,610	51,371
9	115,103	32,988
10 +	293,461	60,316
Total ha	2,631,297	1,679,734

Frequency of spraying is the number of times spray mission flight paths overflow grids in Vietnam GIS developed for herbicide exposure assessment⁴⁶.

Discarded drums

Approximately two litres of herbicide residue remains in the 208-litre barrel after it has been 'emptied'. Typically about 20% of the residue remains after three rinses. Residue, on average, contained 5.96 mg of TCDD, arising from heavily contaminated herbicides and about 1.25 mg per drum from 'low dioxin' herbicides²⁶. Barrel residues had led to inadvertent defoliation of trees and gardens in Da Nang, Nha Trang, Bien Hoa, Phu Cat and Saigon civilian areas near USAF airbases that handled the herbicides when the empty barrels were transported to local merchants for commercial uses³⁵. Improved handling procedures were adopted in 1969 following the Da Nang defoliation incident but the ultimate fate of most of the empty barrels is not known and the extent to which people who used the barrels for other purposes may have been exposed is not known.

Spraying in Laos and Cambodia

Operation Ranch Hand flew its first missions outside RVN in December 1965 to defoliate the major reinforcement and supply route through Laos known as the 'Ho Chi Minh trail'. Known spraying was above the 17th parallel (the northern border of RVN), along Routes 92, 922, 96 and 965 below Tchpone and the Sihanouk Trail that went from Laos to Cambodia. Missions in both countries below the 17th parallel can clearly be seen in Fig. 3a and c. A small amount of crop destruction using Agent Blue was also documented. Laos was also the site of a brief experiment to determine whether F-4E Phantom II jet fighters could successfully be used to carry out spray operations and avoid anti-aircraft fire. At least five F-4E missions were flown until a fighter was shot down and the strategy was abandoned. Documentation of spray activities in Laos is incomplete. The Services-HERBS file shows flight paths for 210 missions, which sprayed about 1.8 million litres. NARA-held documentation shows as much as 14% more herbicides as having been sprayed but no coordinates are given so that these data cannot be included in the revised HERBS file³⁶.

Unlike in Laos, it was official US policy to avoid spraying Cambodia either directly or indirectly by spray drift³⁷ (cited by Cecil³⁸). Records show several heavily sprayed regions of RVN, near Cambodia. The HERBS file shows one five-aircraft Ranch Hand mission dispersing approximately 19,000 l of Agent Orange on 5 April 1969 inside Cambodia. Another nine missions dispersed about 136,000 l of Agent Orange while partly over Cambodian territory (Fig. 3c). At the typical rate of 28 l ha⁻¹ this would cover about 5,500 ha. Undocumented spray drift may have also occurred. In May 1969 a diplomatic crisis arose when Cambodia charged the US with repeatedly spraying it and defoliating 70,930 ha; evidence of defoliation was confirmed by visiting foreign scientists³⁹. Cambodian claims seem to be exaggerated in that to achieve the extent of alleged defoliation nearly half of the Ranch Hand flights for April–May 1969 would have to have been directed towards Cambodia.

Records are not available to resolve the controversy, particularly since the area was devastated by US B-52 bombing raids in 1970.

Discussion

The Vietnam War ended in 1975, yet no large-scale epidemiological study of herbicides and the health of either the Vietnamese population or war veterans has been carried out. Discussions of health and ecology studies in Vietnam have recently taken place⁴⁰. During the course of developing an exposure methodology, we have unexpectedly come upon primary data which expanded existing herbicide spraying databases and could help guide the design of human health and of environmental studies.

NAS-1974 found the HERBS file to be a powerful tool for studying exposure to herbicides. The concordance between HERBS file mission coordinates, operational folders, and the precise locations of roadways, rail lines, power lines, canals and so on given in modern mapping software increases our confidence in the HERBS file. Viewing the HERBS file as a carefully planned target-based military exercise, rather than chronological unrelated missions criss-crossing RVN in straight-line paths affords a coherent analytical approach.

Our analyses using original operational records raise questions about the spraying data and dioxin contamination relied upon by researchers and policymakers. For example, we find that the '202 Tasks Realized' document used by NAS-1974 assumed missions that are missing from the HERBS file⁴¹ to represent targeted areas and anticipated spraying rather than operations completed. (Further, the '202 Tasks Realized' document contains many errors, such as misplaced decimal points for subtotals of volume sprayed, which cumulate and greatly exaggerate the total volume)⁴². Therefore, some NAS-1974 estimates of 'missing' spray must be revised downward. On the other hand, dioxin contamination estimates should be revised upward. Comparatively small amounts of Agent Purple and Pink sprayed in Vietnam between 1961–1965 may have deposited a large percentage of the total dioxin.

Large numbers of Vietnamese civilians appear to have been directly exposed to herbicidal agents, some of which were sprayed at levels at least an order of magnitude greater than for similar US domestic purposes²⁰. Other analyses being carried out by us show large numbers of American troops also to have been directly exposed or to have served in recently sprayed areas. Areas sprayed during the early years and in the various test sites around the world¹⁵ may be of particular interest for follow-up ecological and epidemiological studies. □

Methods

We constructed a new version of the HERBS file using operational records for the USAF 12th Air Commando Squadron (ACS), which carried out the Ranch Hand missions and by systematic comparisons of four different versions of the herbicides files found at the US National Archives⁴³ and the active file at CRUR. Of 18,087 mission records there were 1,264 non-matches between HERB2REV and the current CRUR file (that is, 2.5 million litres out of more than 66.6 million litres). About 3.4% of the total volume is represented at CRUR but not the other files. We also identified errors by mapping flight paths and by consistency checking (for example, legs that were very short or very long, over water, and so on). Discrepancies which we could not thus resolve were reviewed by a small panel of experts. The same panel reviewed each proposed amendment to the HERBS file. A small number of missions which remained ambiguous were discarded. All discarded missions were from the Services-HERBS.

Flight paths for 65 missions of fixed-wing aircraft missions that had been represented in the original HERBS file only by a single point (usually the calculated centre-of-mass) were readily deduced by comparison with similar missions or known targets. (Some missions are correctly represented as single points because they document perimeter spraying carried out by spraying specific discrete points, such as outside base camps near guard posts.) The revised file contains 9,141 missions, primarily by air, but also using other means of delivery.

We used digitizing software⁴⁴ to derive the longitudes and latitudes defining the target perimeters from hand-drawn maps in the USAF Project Folders¹⁹. For folders with no map we extracted the Universal Transverse Mercator (UTM) coordinates from written descriptions of the targets and entered them into a Geographic Information System (GIS) which we created⁴⁵, similar to a databank model created by NAS-1974 (we are willing to assist researchers in using our GIS software for estimating exposure using these data). Population and hamlet exposure estimates used data contained in the HES and Vietnam

gazetteer data files³¹. Each hamlet location was assigned a grid identifier and proximity to spray was calculated using our GIS system software.

Received 6 December 2002; accepted 4 March 2003; doi:10.1038/nature01537.

1. Buckingham, W. A. Jr *Operation Ranch Hand: The Air Force and Herbicides in Southeast Asia 1961–1971* (Office of United States Air Force History, Washington DC, 1982).
2. National Research Council Committee on the Effects of Herbicides in Vietnam. *The Effects of Herbicides in South Vietnam. Part A. Summary and Conclusions* (National Academy of Sciences Press, Washington DC, 1974).
3. Data Management Agency, US MACV. *Herbicide Report System (HERBS)*. Document No. DARU07 (US Military Assistance Command Vietnam, San Francisco, 1970).
4. Heizer, J. R. *Data Quality Analysis of the HERB 01 Data File*. No. MTR-5105 (The Mitre Corporation, Washington DC, 21 April 1971).
5. Wiersma, G. B. *Ecological Impact of Antiplant Agents and Implications for Future Use*. Report No. ACN 16223 (US Army Combat Developments Command, Institute of Land Combat, Fort Belvoir, VA, July 1970).
6. Huddle, F. P. *A Technology Assessment of the Vietnam Defoliant Matter. Report to the Subcommittee on Science, Research and Development, Committee on Science and Astronautics*. (Science Policy Research Division, Legislative Reference Service, Library of Congress Serial F, Washington DC, 8 August 1969).
7. Courtney, K. D. *et al.* Teratogenic evaluation of 2,4,5-T. *Science* **168**, 864–866 (1970).
8. Institute of Medicine Committee to Review the Health Effects in Vietnam Veterans of Exposure to Herbicides. *Veterans and Agent Orange: Health Effects of Herbicides Used in Vietnam* (National Academy of Sciences Press, Washington DC, 1994).
9. Committee on the Assessment of Wartime Exposure to Herbicides in Vietnam, Institute of Medicine. *Characterizing Exposure of Veterans to Agent Orange and Other Herbicides Used in Vietnam: Scientific Considerations Regarding a Request for Proposals for Research* (National Academy of Sciences Press, Washington DC, 1997).
10. Harrigan, E. T. *Calibration Test of the UC-123K/A/A45Y-1 Spray System*. Technical Report No. ADTC-TR-70–36 (Armament and Development Test Center, Eglin AFB Florida, February 1970).
11. Thayer, T. C. *War without Fronts: The American Experience in Vietnam*. Westview Special Studies in Military Affairs (Westview Press, Boulder, CO, 1985).
12. Brown, J. W. *Vegetational Spray Tests in South Vietnam* (US Army Chemical Corps Biological Laboratories, Fort Detrick, MD, 1962).
13. Brown, J. W. *Vegetational Spray Tests in South Vietnam (Supplement U)* (US Army Chemical Corps Biological Laboratories, Fort Detrick, MD, 1962).
14. Olenchuk, P. G., Burke, R. T., Henderson, O. K. & Davis, W. E. *Evaluation of Herbicide Operations in the Republic of Vietnam (September 1962–September 1963)* (US Military Assistance Command, Vietnam, APO San Francisco, 10 October 1963).
15. Darrow, R. A., Truchelat, G. B. & Bartlett, C. M. *OCONUS Defoliation Test Program* (US Army Biological Center, Fort Detrick, MD, July 1966).
16. Cecil, P. F. *Herbicide Warfare. The Ranch Hand Project in Vietnam* (Praeger Scientific, New York, 1986).
17. Collins, C. V. *Herbicide Operations in Southeast Asia July 1961–June 1967*. Report No. DTEC 67–0020 (HQ PACAF, Directorate, Tactical Evaluation, CHECO Division, APO San Francisco, 11 October 1967).
18. Young, A. L., Calcagni, J. A., Thalken, C. E. & Tremblay, J. W. *The Toxicology, Environmental Fate and Human Risk of Herbicide Orange and Its Associated Dioxin*. US AF Technical Report OEHL-TR-78–92 (Brooks AFB, TX, 1978).
19. US Department of Defense. *Records of the US Forces in Southeast Asia, Headquarters, Military Assistance Command Vietnam (MACV), Assistant Chief of Staff for Operations (J3), Chemical Operations Division (MACJ-3–09)*. Herbicide Operations Plans (1966–1967) series, Record Group 472 (National Archives and Records Administration, College Park, MD; 1950–75).
20. Panel on Herbicides of the President's Science Advisory Committee. *Report on 2,4,5-T* (Executive Office of the President, Office of Science and Technology, Washington DC, March 1971).
21. Dwernychuk, L. W. *et al.* Dioxin reservoirs in southern Viet Nam—a legacy of Agent Orange. *Chemosphere* **47**, 117–137 (2002).
22. Young, A. L. *Ecological Studies on a Herbicide-equipment Test Area (TA C-52A)*. (Air Force Armament Laboratory, Eglin AFB, 1974).
23. Letter from Colonel W. T. Moseley to Dr C. E. Minarik. No. MACC007 (US Military Assistance Command Vietnam, APO San Francisco, 10 May 1967).
24. Craig, D. A. *Use of Herbicides in Southeast Asia. Historical Report*. (San Antonio Logistics Center, Directorate of Energy Management, San Antonio AFB, TX, 1975).
25. Darrow, R. A. *Historical, Logistical, Political and Technical Aspects of the Herbicide/Defoliant Program, 1967–1971: A Resume of the Activities of the Subcommittee on Defoliation/Anticrop Systems for JTTCG/CB*. (Vegetation Control Committee, Fort Detrick, MD, September 1971).
26. *Revised Draft Environmental Impact Statement. Disposition of Orange Herbicide by Incineration*. Report No. AF-ES-72–2D. (Department of the Air Force, Washington, DC, April 1974).
27. Hughes, B. M. *et al.* *Analytical Methodology for Herbicide Orange. Vol 1: Determination of Chemical Composition*. No. ARL 75–0110 (Wright-Patterson AFB, Aerospace Research Laboratories, OH, May 1975).
28. Fee, D. C., Hughes, B. M., Tiernan, T. O., Hill, C. E. & Taylor, M. L. *Analytical Methodology for Herbicide Orange. Vol II: Determination of Origin of USAF Stocks*. No. ARL 75–0110 (Wright-Patterson AFB, Aerospace Research Laboratories, OH, May 1975).
29. Darrow, R. A., Irish, K. R. & Minarik, C. E. *Herbicides Used in Southeast Asia*. (United States Army Plant Sciences Laboratories, Fort Detrick, Maryland, 1969).
30. Young, A. L. & Reggiani, G. M. *Agent Orange and Its Associated Dioxin: Assessment of a Controversy* (Elsevier, Amsterdam, 1988).
31. US Department of Defense. *Hamlet Evaluation System (HES)*. [Electronic Records]. Record Group 330 (National Archives and Records Administration Center for Electronic Records, College Park, MD, 1967–74).
32. Orians, G. H. & Pfeiffer, E. W. United States goals in Vietnam. *Science* **169**, 1030 (1970).
33. Smith, C. & Watkins, D. *The Vietnam Map Book. Self-help Guide to Herbicides Exposure*. (Winter Soldier Archive, Berkeley, CA, November 1981).
34. Blackman, G. E., Fryer, J. D., Lang, A. & Newton, M. *The Effects of Herbicides in South Vietnam. Part B:*

- Working Papers. Persistence and Disappearance of Herbicides in Tropical Soils* (National Academy of Sciences-National Research Council, Washington DC, 1974).
35. US Military Assistance Command, Vietnam. *Vietnam Lessons Learned No. 74: Accidental Herbicide Damage* (MACV, APO San Francisco, 15 September 1969).
 36. US Department of Defense *Records of the US Forces in Southeast Asia, 1950–1975, Headquarters, Military Assistance Command Vietnam (MACV), Assistant Chief of Staff for Operations (J3) Record Group 472* (National Archives and Records Administration, Modern Military Records LICON, Textual Archives Services Division, College Park, MD, 1950–75).
 37. American Embassy, Phnom Penh. Telegram No. 875, NSF-VN Vol. X, Box 196–98 (John F. Kennedy Presidential Library, Boston, MA, 18 May 1963).
 38. Cecil, P. F. *The Air Force Ranch Hand Project in Southeast Asia: Operations and Consequences* PhD dissertation, Texas A&M Univ. (1984).
 39. Hay, A. *The Chemical Scythe: Lessons of 2,4,5-T and Dioxin 187–195* (Plenum, New York, 1982).
 40. Cyranoski, D. US and Vietnam join forces to count cost of Agent Orange. *Nature* **416**, 262 (2002).
 41. '202 Committee' *List of 202 Task (sic) Realized From January 1962 to September 1965*. (Military Assistance Command, Vietnam, Saigon, undated).
 42. Dashiell, T. R. (Office of the Director of Defense Research and Engineering) Comments to *Dr. P. Kunstader's Comparison of HERBS and 202 Records* (Memorandum to P. Ross, National Academy of Sciences, 11 July 1973).
 43. US Department of Defense *Herbicide Files, 1965–1970*. Record Group 30 (National Archives and Records Administration Center for Electronic Records, College Park, MD, 1965–70).
 44. R2V Software Able Software Corp, 2000; www.ablesw.com.
 45. Stellman, J. M. *et al.* A Geographic Information System for characterizing exposure to Agent Orange and other herbicides in Vietnam. *Environ. Health Perspect* **111**, 321–323 (2003).
- Acknowledgements** This research was supported by the National Academy of Sciences, the Department of Veterans Affairs and USPHS at the National Cancer Institute. We thank D. Hakenson and his staff at the US Armed Services Center for Research of Unit Records for technical support and access to military records; the staff of the National Archives and Records Administration, particularly R. Boylan and M. O. Adams; N. Heim for assistance with illustrations and F. Benjamin for assistance with military records. R.C. is a retired member of the US Army.
- Competing interests statement** The authors declare that they have no competing financial interests.
- Correspondence** and requests for materials should be addressed to J.M.S. (e-mail: jms13@columbia.edu).

Injection of adult neurospheres induces recovery in a chronic model of multiple sclerosis

Stefano Pluchino*, Angelo Quattrini†‡, Elena Brambilla*, Angela Gritti§, Giuliana Salani*, Giorgia Dina†, Rossella Galli§, Ubaldo Del Carro‡, Stefano Amadio‡, Alessandra Bergami*, Roberto Furlan*‡, Giancarlo Comi‡, Angelo L. Vescovi§ & Gianvito Martino*‡

* Neuroimmunology Unit—DIBIT, † Neuropathology Unit, § Stem Cell Research Institute, and ‡ Department of Neurology and Neurophysiology, San Raffaele Hospital, via Olgettina 58, 20132 Milano, Italy

Widespread demyelination and axonal loss are the pathological hallmarks of multiple sclerosis. The multifocal nature of this chronic inflammatory disease of the central nervous system complicates cellular therapy and puts emphasis on both the donor cell origin and the route of cell transplantation. We established syngenic adult neural stem cell cultures and injected them into an animal model of multiple sclerosis—experimental autoimmune encephalomyelitis (EAE) in the mouse—either intravenously or intracerebroventricularly. In both cases, significant numbers of donor cells entered into demyelinating areas of the central nervous system and differentiated into mature brain cells. Within these areas, oligodendrocyte progenitors markedly increased, with many of them being of donor origin and actively remyelinating axons. Furthermore, a significant reduction of astrogliosis and a marked decrease in the extent of demyelination and axonal loss were observed in transplanted animals. The functional impairment caused by EAE was almost abolished in transplanted mice, both clinically and neurophysiologically. Thus, adult neural precursor cells promote multifocal remyelination and functional recovery after intravenous or intrathecal injection in a chronic model of multiple sclerosis.

The permanent neurological impairment typical of chronic inflammatory demyelinating disorders of the central nervous system (CNS), such as multiple sclerosis, is due to the axonal loss resulting from recurrent episodes of immune-mediated demyelination^{1,2}. So far, experimental cell therapy for these disorders has been based mainly on the transplantation of myelin-forming cells, or their precursors, at the site of demyelination^{3–6}. Although such an approach can trigger functional recovery and restore axonal conduction, the limited migration of lineage-restricted, myelin-forming cells through the brain parenchyma highlights the beneficial effect of transplantation to the site of the injury^{7,8}. This raises critical issues regarding the therapeutic use of focal cell transplantation to treat diseases in which multifocal demyelination is the main pathological feature. Such issues are compounded by the poor expansion capacity of myelin-forming cells in culture^{7,8}, which greatly limits their availability and further hampers their prospective application in clinical settings.

Multipotent neural precursors—with the capacity to generate neurons, astroglia and oligodendroglia—are found in the adult brain and possess the critical features of somatic stem cells⁹. They support neurogenesis within restricted areas throughout adulthood, can undergo extensive *in vitro* expansion and, therefore, have been proposed as a renewable source of neural precursors for regenerative transplantation in various CNS diseases^{10,11}.

Here, we investigate the ability of these neural precursors to reach multiple demyelinating areas of the CNS and to ameliorate disease-related, clinical, neurophysiological and pathological signs when injected both intraventricularly or systemically in an experimental model of multiple sclerosis—EAE. Throughout this study, we used cultures of neural stem cells that were established from the periventricular region of the forebrain ventricles of adult C57BL/6 mice, expanded *in vitro*¹², and labelled with a lentiviral vector by which the expression of the *Escherichia coli* β -galactosidase (β -gal) gene (*lacZ*) was directed to the nuclear compartment. Clonal, subcloning and population studies confirmed the functional stability and stem cell composition of these neural precursor preparations (Supplementary Fig. A), as previously described¹³.

Migration and homing

In CNS inflammation, the upregulation of surface cell adhesion molecules in peripheral blood cells is a prerequisite for their interaction with activated endymal¹⁴ and endothelial cells^{6,15–17} and their crossing of the blood–brain barrier. Thus, we investigated surface adhesion molecule expression in our cultures. Most (>90%) of the cells expressed CD44 and very late antigen (VLA)-4, but not P-selectin glycoprotein ligand (PSGL)-1, L-selectin and leukocyte function-associated antigen (LFA)-1 (Supplementary Fig. A), indicating that the donor cells in our study might have the capacity to actively interact with the blood–brain barrier and to enter the CNS parenchyma.

To investigate the migratory capacities of neural precursors *in vivo*, we injected them into the cisterna magna (intracerebroventricularly, i.c.) or into the blood stream (intravenously, i.v.) of syngenic, untreated C57BL/6 mice as well as in mice that had been pretreated with lipopolysaccharide (LPS; i.v.) or with tumour-necrosis factor- α (TNF- α) and interleukin (IL)-1 β ¹⁷ (i.c.). Both procedures were used to force the opening of the blood–brain barrier, thus mimicking a CNS-restricted inflammatory-like condition similar to that occurring during the early phases of EAE or multiple sclerosis. One day after transplantation, none of the i.v.- or i.c.-injected cells could be detected within the CNS of untreated mice (data not shown). In contrast, numerous donor cells were found within meningeal spaces (Fig. 1a, b) or around post-capillary venules (Fig. 1g, h) in mice pretreated with LPS or TNF- α /IL-1 β , depending on whether neural precursors were injected i.c. or i.v. This finding indicated that donor cells might have the ability to pass through the blood–brain barrier and enter into the CNS under inflammatory conditions.

Next, we expanded our investigation by injecting neural precursors (1×10^6 cells per mouse) into C57BL/6 mice affected by myelin oligodendrocyte glycoprotein (residues 35–55) (MOG(35–55))-induced chronic EAE¹⁸. Neural precursors were transplanted through i.c. or i.v. injection, either before disease onset (for example, 10 days post-immunization (p.i.)), at the onset of the disease (day 15 p.i.), or one week later (day 22 p.i.). Ten days after

transplantation, the donor cells were distributed along the whole anterior–posterior neural axis (neuraxis) exclusively in those mice with EAE that received neural precursors through i.c. or i.v. injection after the onset of the disease. Notably, although i.c.-injected cells were localized mainly within the submeningeal space in close proximity to subpial inflammatory foci (Fig. 1c–f), the i.v.-injected cells were found around post-capillary venules (Fig. 1i), almost exclusively in areas of CNS damage (Fig. 1j, k). At 10 days after transplantation, relatively few cells were found in the CNS parenchyma. Thirty days after injection of neural precursors, many of the donor cells were localized deeply within the brain parenchyma and displayed a marked distribution pattern: most of them were confined within areas of demyelination and axonal loss, and only very few were found in regions within which the myelin architecture was preserved (Fig. 2).

Engraftment and differentiation

Detailed examination of tissues showed that cells derived from the neural precursors were located closely proximal to demyelinated axons or to axons undergoing a process of active remyelination (Fig. 2a–d). Some of these cells were morphologically indistinguishable from endogenous oligodendrocyte progenitors, with their cell membrane being in close contiguity with the myelin sheets of

some of the surrounding remyelinating axons (Fig. 2a–k). As oligodendrocyte progenitors are able to complete a full programme of myelin re-ensheathment¹⁹, the ensuing idea that neural precursors may differentiate into myelin-forming cells was unequivocally confirmed by electron microscopy analysis. Transplanted cells, showing perinuclear β -gal deposits (Fig. 2e, g, j), were found in close contact with axons that were surrounded by poorly compacted myelin sheets (Fig. 2f, i), a well-established morphological hallmark of active remyelination (Supplementary Fig. B).

Neuropathological analysis of the CNS of EAE mice injected with neural precursors revealed an average density of 19.5 ± 1.1 and 22.1 ± 1.1 β -gal⁺ cells mm⁻² in the brain and 8.6 ± 0.9 and 7.0 ± 1.0 β -gal⁺ cells mm⁻² in the spinal cord of i.c.- and i.v.-injected mice, respectively. By means of lineage-specific neural markers, we established that a consistent part of the engrafted cells differentiated into platelet-derived growth factor- α (PDGF α) receptor-expressing oligodendrocyte precursors (i.c.-injected mice: 8.1 ± 1.1 and 3.7 ± 0.9 cells mm⁻² for brain and spinal cord, respectively; i.v.-injected mice: 9.4 ± 1.0 and 5.3 ± 0.7 cells mm⁻² for brain and spinal cord, respectively) (Fig. 3a, f). The remaining injected cells generated astrocytes (i.c.-injected mice: 1.9 ± 0.4 and 0.2 ± 0.1 cells mm⁻² for brain and spinal cord; i.v.-injected mice: 2.9 ± 0.5 and 0.1 ± 0.1 cells mm⁻² for brain and spinal cord)

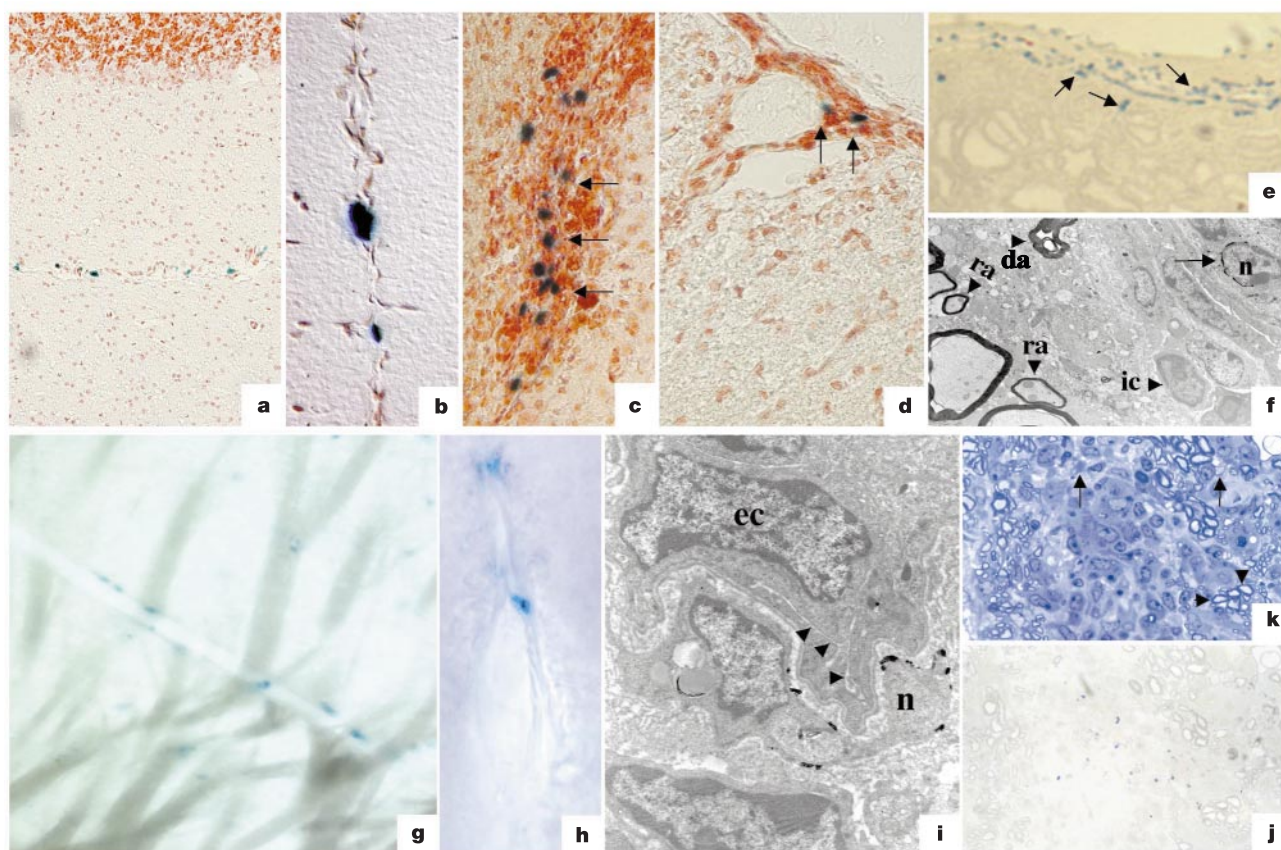


Figure 1 Distribution of *nls-lacZ*-labelled, syngenic neural precursors after i.c. (a–f) and i.v. injection (g–k) in C57BL/6 mice. One day after injection in mice, in which the blood–brain barrier was forced open using pro-inflammatory cytokines¹⁷ (a, b) or LPS¹⁸ (g, h), i.c.-injected neural precursors (blue cells) were located within the meningeal space (a, cerebellum, $\times 20$; b, main sagittal superior scissure, $\times 100$), whereas i.v.-injected neural precursors were confined primarily to perivascular or intravascular areas (g, deep white matter of the lateral ventricles of the brain, $\times 40$; h, white matter of the spinal cord, $\times 40$). Ten days after i.c. injection in EAE mice, neural precursors (arrows) were found both in the brain (c, cerebellum, $\times 40$) and in the spinal cord (d, posterior column, $\times 40$). Furthermore, they were located mainly in the submeningeal space (e, arrows), close to

areas infiltrated by inflammatory cells (ic) containing both remyelinating (ra) and demyelinated axons (f, da, $\times 1,500$). In turn, i.v.-injected neural precursors were found within areas in which clusters of demyelinated (k, arrows, $\times 250$) and normal-appearing axons (k, arrowheads) were visible. i, Neural precursors located across the blood–brain barrier and in close contact with endothelial cells (ec) of the blood–brain barrier were also visible (vascular lumen, arrowheads). Neural precursors are identified by light microscopy (e, $\times 600$; j, $\times 250$) and by electron microscopy (f, $\times 1,500$; i, $\times 6,500$) as cells whose nuclei (n) are surrounded by blue deposits or by rod-shaped electron-dense precipitates (arrows), respectively.

(Fig. 3c) and neurons (i.c.-injected mice: 5.2 ± 0.8 and 2.8 ± 0.6 cells mm^{-2} for brain and spinal cord; i.v.-injected mice: 7.2 ± 1.0 and 4.4 ± 0.7 cells mm^{-2} for brain and spinal cord) (Fig. 3d, e) or retained undifferentiated nestin⁺ morphological features (i.c.-injected mice: 3.0 ± 0.4 and 0.3 ± 0.1 cells mm^{-2}

for brain and spinal cord; i.v.-injected mice: 4.0 ± 0.8 and 0.1 ± 0.1 cells mm^{-2} for brain and spinal cord) (Fig. 3b). Notably, in both i.c.- and i.v.-injected mice, the number of astrocytes derived from neural precursors and that of cells retaining an undifferentiated morphology decreased significantly ($P < 0.005$) in the spinal cord (1.3% astrocytes and 2.5% undifferentiated), as compared with the brain (11.5% astrocytes and 16.8% undifferentiated). Conversely, the number of neural-precursor-derived oligodendrocyte precursors and that of neurons increased in the spinal cord (50.6% oligodendrocyte precursors and 45.6% neurons), as compared with the brain (41.8% oligodendrocyte precursors and 29.8% neurons). A detailed exploration of tissues other than the CNS showed that i.c.- or i.v.-injected neural precursors also reached many of the bodily organs (such as lung, liver, spleen and kidney) within 10 days after transplantation, but could no longer be detected 20 days later (Supplementary Fig. C).

Modulation of growth factor expression

We observed notable changes within the demyelinating areas into which neural precursors had migrated. The typical glia scar surrounding these lesions was consistently and visibly reduced, as indicated by the significant lower density of interleaving astrocyte processes (suggestive of reactive astrogliosis) in demyelinating areas of mice treated i.c. (Fig. 4a) or i.v. (Fig. 4b) with neural precursors, as compared with controls (Fig. 4c). Accordingly, we detected a concomitant decrease in the CNS levels of messenger RNA species

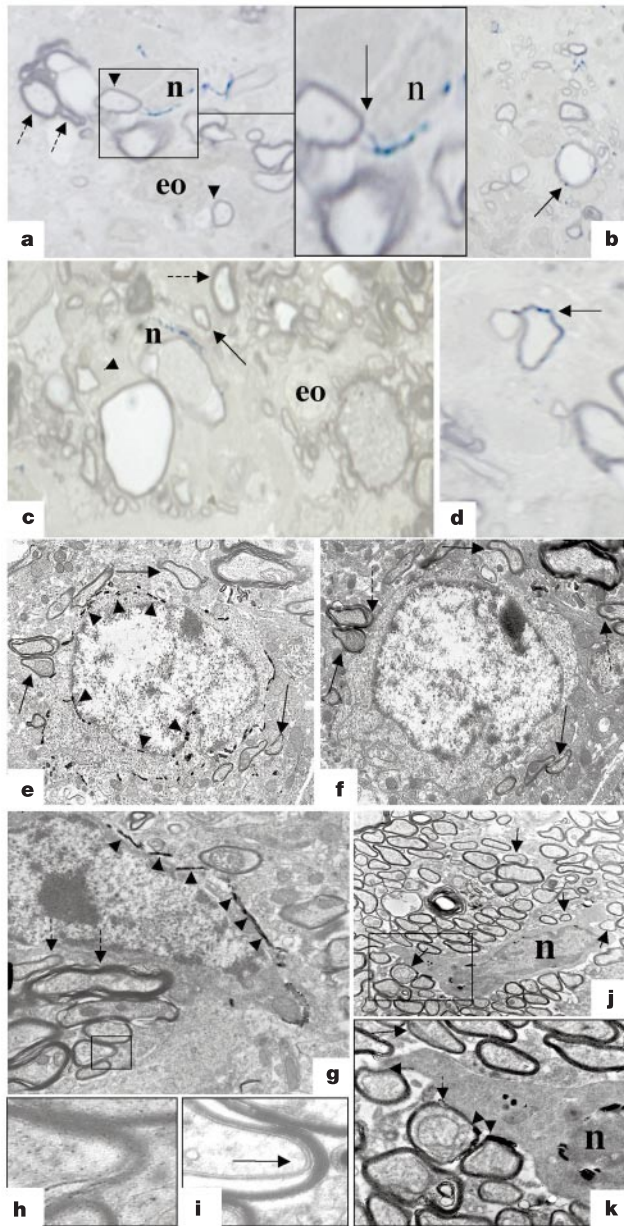


Figure 2 Neural precursors contribute to remyelination of demyelinated axons in EAE mice. **a–d**, Light microscopy showing transplanted neural precursors, the nuclei (n) of which are surrounded by blue deposits, entering into close contact (arrows) with axons surrounded by a larger rim of myelin (dashed arrows) or by a thin rim of myelin (arrowheads). In the same areas, endogenous oligodendrocyte-like cells (eo) coexist. Blue deposits are visible close to myelin lamellae (arrows) (**a**, **b** and **d**, $\times 600$; **c**, $\times 800$). **e–k**, Electron microscopy showing i.v.-injected β -gal⁺ neural precursors (arrowheads) entering into close contact with various axons and interdigitating within myelin fibres. The cytoplasmic membrane of β -gal⁺ neural precursors contacts several axons surrounded by a thin rim of myelin (arrows). β -gal⁺ dark deposits are also seen close to myelin lamellae (dashed arrows) (**e** and **f**, $\times 3,000$). **g**, **j**, Axons, which are in close contact with the cell membrane of β -gal⁺ cells (**i**; arrow), are surrounded by poorly compacted, newly formed myelin. **h**, **i**, **k**, High-magnification of boxed areas in **g** and **j** (**h** and **i**, $\times 18,000$; **k**, $\times 4,400$).

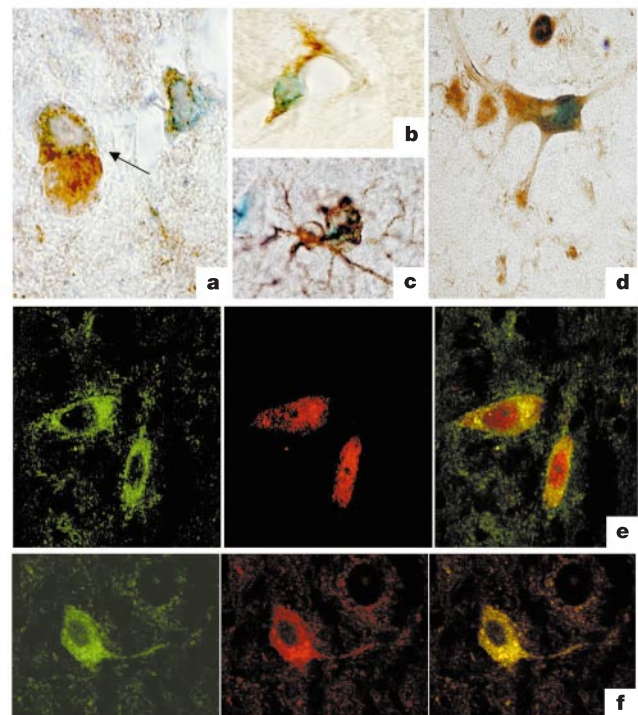


Figure 3 Differentiation of engrafted neural precursors into mature neural cells in EAE mice. **a–d**, Neural-precursor-derived cells (stained blue), which were co-labelled with the oligodendrocyte precursor marker PDGF α receptor, were found within the CNS (arrow) of EAE mice that received either an i.c. or an i.v. injection of neural precursors (**a**, $\times 100$). Within the same area, endogenous unlabelled progenitors were also detected. Blue cells, which were double-labelled with the undifferentiated cell marker nestin (**b**, $\times 100$), the astroglia marker GFAP (**c**, $\times 100$) or the neuronal marker NeuN (**d**, $\times 100$), were also detected. **e**, **f**, Confocal microscopy showing co-localization of eGFP (green) and the neuronal marker NeuN (**e**, $\times 100$; red) or the oligodendrocyte precursor marker PDGF α receptor (**f**, $\times 100$; red). Engraftment and differentiation of eGFP⁺ neural precursors were comparable to those of β -gal⁺ neural precursors (data not shown).

coding for factors known to drive reactive astrogliosis, such as fibroblast growth factor (FGF)-II²⁰ ($P < 0.05$) and transforming growth factor (TGF)- β ²¹ ($P = 0.06$). Expression levels of mRNA of other neurotrophic growth factors, such as ciliary neurotrophic factor (CNTF), neurotrophin (NT)-3, nerve growth factor (NGF), glial-derived neurotrophic factor (GDNF), brain-derived neurotrophic factor (BDNF) and leukaemia inhibitory factor (LIF), remained unchanged (Fig. 4d). Finally, the overall density of oligodendrocyte progenitors was significantly ($P < 0.001$) increased within demyelinating areas of the spinal cord of mice injected i.c. (27.9 ± 3.9 cells mm^{-2}) (Fig. 4e) or i.v. (25.2 ± 2.3 cells mm^{-2}) (Fig. 4f) with neural precursors, as compared with controls (5.1 ± 0.5 cells mm^{-2}) (Fig. 4g).

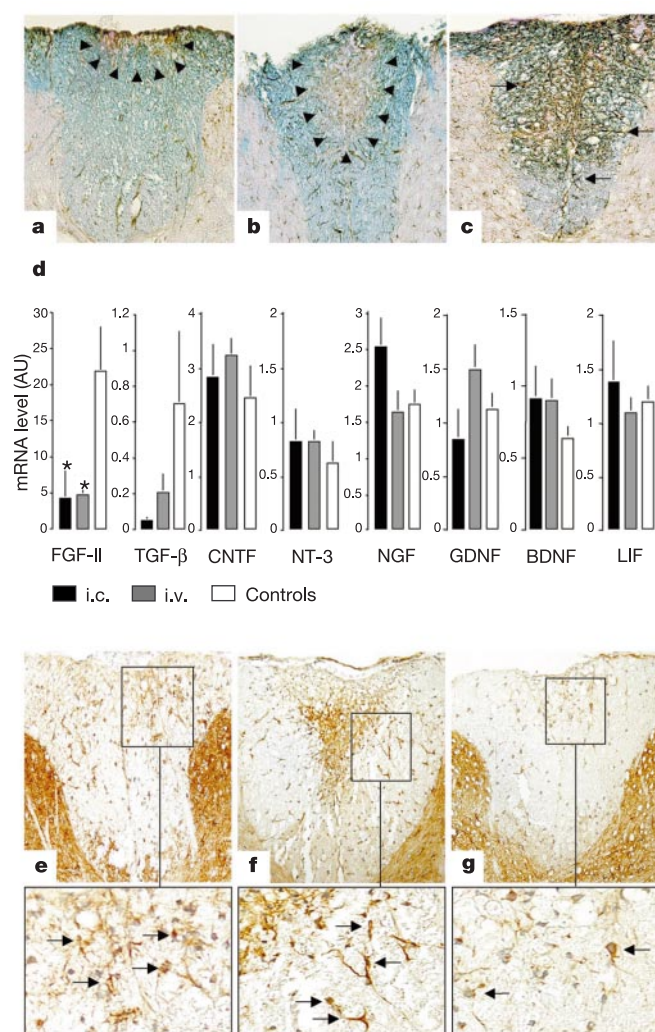


Figure 4 I.c. or i.v. injection of neural precursors into EAE mice reduces glia scarring and modulates neurotrophic growth factor mRNA expression within the CNS. **a–c**, Reduction of reactive astrogliosis occurring within demyelinating areas (arrowheads) in mice injected i.c. (**a**) or i.v. (**b**) with neural precursors, as compared with sham-treated controls (**c**). The blue labelling by Luxol fast staining shows the intact myelin. Interleaving astrocyte processes (arrows; GFAP staining), indicative of reactive astrogliosis, were found in sham-treated but not in neural-precursor-transplanted mice. **d**, mRNA levels of neurotrophic factors in the CNS of sham-treated mice (open bars) and mice injected i.c. (black bars) or i.v. (grey bars) with neural precursors. FGF-II mRNA was significantly (asterisk, $P < 0.05$) reduced in transplanted mice. **e–g**, Increase in PDGF α receptor-expressing oligodendrocyte progenitors (brown stain) within the posterior columns of the spinal cord of EAE mice transplanted with neural precursors either by i.c. (**e**) or i.v. injection (**f**), as compared with sham-treated animals (**g**). Magnification in **a–c** and **e–g**, $\times 20$.

Protection from disease progression and recovery

Given the homing of neural precursors into the CNS of EAE mice and the considerable histopathological improvement, we investigated whether their transplantation by the i.c. or i.v. route would also result in clinical recovery in these mice. One week after the onset of the disease (on day 22 p.i.), EAE mice were injected i.c. or i.v. with PBS (hereafter referred to as sham-treated mice), or with equal numbers (1×10^6 cells per mouse) of unsorted syngenic whole bone marrow cells, syngenic fibroblasts or syngenic neural precursors. Moreover, to assess whether transplanted neural precursors might interfere with the ongoing MOG-directed immune response, either working as decoy antigens or by desensitization, dead neural precursors were also injected i.v. into a subgroup of EAE mice. Prominent clinical amelioration was observed exclusively in mice receiving viable neural precursors. Mice receiving neural precursors i.c. began recovering 15 days after cell transplantation (Fig. 5a), whereas i.v.-transplanted mice recovered faster (Fig. 5b), starting 5 days earlier (10 days after cell injection). At the end of the follow-up period (45 days p.i.), both i.c.- and i.v.-treated mice showed a significant recovery from the EAE-related clinical deficits ($P \leq 0.005$ and $P \leq 0.01$, respectively). From a practical perspective, although sham-treated EAE mice were affected by an overt paresis of the hindlimbs, those transplanted with neural precursors displayed clinical features that fell somewhere between the normal behaviour (full recovery) and a minor tail paralysis. Moreover, among the EAE mice completing the follow-up period, 3 of 11 (27%) of those injected i.c. and 4 of 15 (26.6%) of those injected i.v. with neural precursors recovered completely from the disease ($P < 0.05$). None of the mice that received cell types other than neural precursors showed any sign of recovery compared with sham-treated mice (Fig. 5c–e). Clinical recovery was accompanied by a significant decrease in the extent of demyelination ($P < 0.001$) and in axonal loss ($P < 0.01$) 30 days after cell injection (Table 1).

To demonstrate further the functional recovery triggered by injection of neural precursors, we measured central conduction time (CCT) by means of motor-evoked potentials (MEPs) in sham-treated controls and EAE mice treated with neural precursors. CCT is a measure of the time of propagation of the electrical stimulus from motor cortex to spinal motor neurons, with a prolonged or no measurable CCT, indicating slower motor pathways as a consequence of demyelination or axonal damage. Forty-five days p.i., CCT was measurable in all mice, but was significantly ($P \leq 0.05$) reduced in i.c.- (2.97 ± 0.14 ms) and i.v.-treated mice (3.12 ± 0.14 ms) compared with sham-treated controls (3.59 ± 0.18 ms) (Fig. 5f). This positive effect was long lasting. Eighty days after cell transplantation, CCT was measurable in all of the mice injected i.v. (6 mice) or i.c. with neural precursors (9 mice), but in only 4 of 8 (50%) of the sham-treated controls. CCT was closer to normal levels in i.c.- (3.41 ± 0.41 ms) and i.v.-treated mice (3.69 ± 0.49 ms), but it was significantly ($P \leq 0.025$) prolonged in sham-treated controls (4.58 ± 0.89 ms). Doubling either the amount of cells transplanted in each mouse or the number of injections did not result in any increase or quickening in the observed beneficial effects compared with standard injection (1×10^6 cells per mouse) (Supplementary Fig. D).

In vitro modulatory properties

Clinical amelioration in mice injected with neural precursors began 10–15 days after cell transplantation in EAE mice. Because during this time neural precursors and their progeny underwent differentiation, we investigated the basal expression of mRNA species encoding neurotrophic factors in neural precursors and the possible changes that may occur when these cells give rise to their mature neuronal/glia progeny¹². Detectable levels of mRNA for FGF-II, TGF- β , CNTF, GDNF, NGF, BDNF, LIF and NT-3 (Supplementary Fig. E) were found in cultured neural precursors, which were

Table 1 EAE features in C57BL/6 mice treated i.c. or i.v. with neural precursors

Treatment	Route of cell administration	No. of mice	Disease onset (days p.i.)	Maximum clinical score	Cumulative disease score (0–22 p.i.)*	Cumulative disease score (23–45 p.i.)	Inflammatory infiltrates (no. per mm ²)†	Demyelination (% mm ⁻²)‡	Axonal loss (% mm ⁻²)‡
Sham-treated	–	31	14.4 ± 0.6	2.8 ± 0.1	20.5 ± 1.6	49.3 ± 2.1	5.1 ± 0.4	2.6 ± 0.3	3.6 ± 0.5
Neural precursors	i.c.	17	14.1 ± 0.3	2.5 ± 0.1	16.7 ± 1.5	38.4 ± 3.3‡	4.5 ± 0.8	0.7 ± 0.2	1.4 ± 0.3#
Neural precursors	i.v.	22	14.2 ± 0.5	2.6 ± 0.1	17.9 ± 1.4	36.9 ± 3.1§	4.8 ± 0.6	0.6 ± 0.1¶	1.4 ± 0.3☆
Whole bone marrow	i.v.	5	14.0 ± 0.7	2.9 ± 0.3	18.1 ± 1.8	51.9 ± 9.2	6.1 ± 0.5	1.7 ± 0.3	3.8 ± 1.1
Fibroblasts	i.v.	5	15.6 ± 0.6	3.1 ± 0.3	14.1 ± 2.9	54.0 ± 5.6	4.8 ± 0.3	4.1 ± 0.4	4.2 ± 0.6

Data are means ± standard error. i.c., intracerebroventricularly; i.v., intravenously; p.i., post-immunization; whole bone marrow, unsorted whole bone marrow cells.

*The cumulative score represents the summation of the single score recorded in each mouse before (from the day of immunization (day 0) to day of cell transplantation (day 22)) and after (from day 23 to day 45 p.i.) cell transplantation.

†Inflammatory infiltrates, demyelination and axonal loss have been quantified on an average of 12 spinal cord sections per mouse for a total of five mice per group.

‡ $P < 0.01$ when compared with sham-treated controls.

§ $P < 0.005$ when compared with sham-treated controls.

|| $P < 0.0001$ when compared with sham-treated controls; $P < 0.0001$ and $P < 0.05$ when compared with fibroblasts injected i.v. and whole bone marrow cells injected i.v., respectively.

¶ $P < 0.0001$ when compared with sham-treated controls; $P < 0.0001$ and $P < 0.01$ when compared with fibroblasts injected i.v. and whole bone marrow cells injected i.v., respectively.

$P = 0.001$ when compared with sham-treated controls; $P < 0.001$ and $P < 0.05$ when compared with fibroblasts injected i.v. and whole bone marrow cells injected i.v., respectively.

☆ $P < 0.01$ when compared with sham-treated controls; $P < 0.01$ when compared with fibroblasts injected i.v.

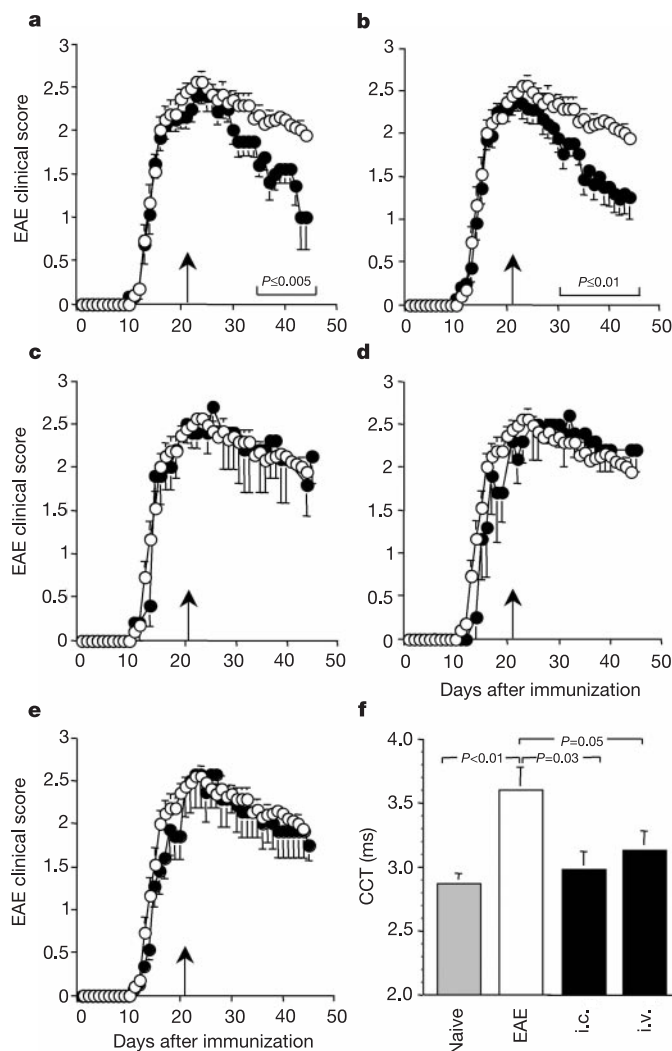


Figure 5 i.v. and i.c. injection of neural precursors after disease onset (arrow) significantly improves clinical features in EAE mice. EAE clinical score (see Methods) in mice injected with different cell types (filled circles) and in sham-treated mice (open circles; $n = 31$). Only the mice that received neural precursors i.c. (**a**, $n = 17$) or i.v. (**b**, $n = 22$) display a pronounced clinical improvement compared with sham-treated mice and mice injected i.v. with unsorted whole bone marrow cells (**c**, $n = 5$), fibroblasts (**d**, $n = 5$) or dead neural precursors (**e**, $n = 7$). **f**, Neurophysiological assessment of myelin conduction

velocity in EAE mice transplanted with neural precursors (i.v., $n = 9$; i.c., $n = 5$), sham-treated EAE mice ($n = 11$) and naive untreated mice ($n = 14$). Significantly ($P \leq 0.05$) lower CCT values, showing improvement of myelin conduction velocity, were recorded in neural-precursor-treated mice when compared with sham-treated animals. Data in the graph are expressed as means ($\pm$ standard error) of at least five different experiments.

unaffected after their differentiation *in vitro*. Notably, CNTF—a factor critically involved in protection from CNS demyelination²²—was fivefold higher in differentiated cells than in neural precursors (Supplementary Fig. E). As the neuroprotective effect of CNTF in EAE²² has been attributed to *in situ* downregulation of TNF- α , we also measured the levels of mRNA for TNF- α both in neural precursors before and after their differentiation *in vitro*, and in the CNS of i.c.- and i.v.-treated EAE mice. Undifferentiated and differentiated neural precursors did not appear to produce TNF- α , which, however, was significantly downregulated in the CNS of mice treated with neural precursors, as compared with controls (Supplementary Fig. E). Accordingly, we also found that the mRNA levels of matrix metalloproteinases (MMP)-2 and -9—two molecules that act synergistically with TNF- α in the breaking of the blood–brain barrier and the facilitation of CNS transport and migration of encephalitogenic lymphocytes in EAE/multiple sclerosis²³—were significantly lower in EAE mice treated with neural precursors, as compared with untreated mice, although the genes *Mmp2* and *Mmp9* were not transcribed in undifferentiated neural precursors and their mature progeny *in vitro* (Supplementary Fig. E).

Bimodal mechanism of action

Our results show that the mechanism underlying the neural-precursor-mediated clinical amelioration is bimodal. First, neural precursors differentiate into myelin-forming cells and establish a programme of anatomical and functional re-ensheathing of myelin at the site of the lesion. Second, in demyelinated areas of mice receiving neural precursor injection, oligodendrocyte progenitors increase by fivefold compared with those in sham-treated controls. Yet, although up to 20% of these derive from donor neural precursors, the remainder are of endogenous origin. Furthermore, astrogliosis—which hampers endogenous axonal remyelination, leading to irreversible axonal loss and functional disability—is greatly reduced in animals transplanted with neural precursors. This shows that, in addition to carrying out direct remyelination, transplanted neural precursors can also act as bystander regulators of endogenous oligodendroglia and of reactive astrogliosis in EAE. Some of the effectors of this ‘humoral’ mechanism may perhaps be molecules that are critically involved in this phenomenon, such as FGF-II and TGF- β ^{20,21}, whose expression is modulated in the CNS after neural precursor transplantation. Nonetheless, it is likely that many other molecules—such as pro-inflammatory cytokines, growth factors (for example, CNTF)²² and/or MMPs²³ that are involved in the effector’s phase of autoimmune demyelination—may also be implicated in this phenomenon.

Cell therapy is a prominent area of investigation in the biomedical field, particularly for the treatment of otherwise incurable neurodegenerative CNS disorders. In this view, both embryonic and neural stem cells are being proposed as an elective source of brain cells for transplantation²⁴. Our findings show that cultures of neural stem cells may function as a therapeutic tool to improve clinicopathological signs and symptoms of chronic inflammatory demyelinating diseases of the adult brain. This reinforces the concept that somatic stem cells may represent an effective ‘weapon’ for the cure of CNS disorders where neurodegeneration causes permanent disability. These findings acquire particular relevance when considering that the pathology studied here is a model of multiple sclerosis, a CNS demyelinating disease of a complex type, presenting us with a particularly difficult challenge. The problems posed by the chronic and inflammatory nature of the disease—through the formation of glia scars and by the depletion of the local pool(s) of oligodendrocyte progenitors that should re-ensheath demyelinated axons^{19,25}—are greatly compounded by the presence of multiple demyelinating plaques diffused throughout the neuraxis. This pathological feature prevents the use of a classical neural transplantation approach, in which the therapeutic cells are injected in the

proximity of the site of the lesion or into its target area, such as in Parkinson’s disease²⁶.

Somatic stem cells can target injured CNS tissue and promote functional recovery^{27,28}. However, cells may promote functional recovery in multiple, diffused areas of the CNS only when injected *in situ* during fetal life²⁹, whereas in adulthood their beneficial effect(s) seem to be achievable only in focal injuries^{11,27–30}. To the best of our knowledge, this work provides the initial evidence showing how neural precursors may represent a renewable source of cells which, when transplanted into the cerebroventricular system or into the blood stream, can reach multiple areas of a chronically injured adult CNS, enter the brain tissue and seek damaged areas where they promote structural and functional recovery. The possibility of injecting therapeutic cells systemically to achieve significant clinical benefit in multiple sclerosis-like syndromes opens new opportunities for the clinical use of stem-cell-based therapies to treat heretofore incurable diseases in humans. □

Methods

Cell preparation and clonogenic analysis

Neural precursor cultures were established and expanded as previously described (see also Supplementary Information)¹². Neural precursors from passage 5 to 10 were used throughout this study. Cells were labelled *in vitro* using a third-generation lentiviral vector pRRLsin.PPT-hCMV engineered with the *E. coli*-derived *lacZ* gene containing a nuclear localization signal (*nls*)³¹. We labelled more than 80% of the cells by this method. Clonal and population studies confirmed the functional stability of the cells used here, as shown previously (see also Supplementary Information)¹³. Whole bone marrow unsorted cells were obtained by flushing femurs and tibiae of 6–8-week-old C57BL/6 mice. Fibroblast cultures were prepared from adult C57BL/6 kidneys dispersed through a 70- μ m cell strainer and maintained in DMEM supplemented with 10% FCS and antibiotics.

EAE induction and treatment

Chronic, relapsing EAE was induced in 122 C57BL/6 mice by subcutaneous immunization with 300 μ l of 200 μ g MOG(35–55) (Multiple Peptide System) in incomplete Freund’s adjuvant containing 8 mg ml^{−1} *Mycobacterium tuberculosis* (strain H37Ra; Difco). Pertussis toxin (Sigma) (500 ng) was injected on the day of the immunization and again two days later. Body weight and clinical score (0 = healthy; 1 = limp tail; 2 = ataxia and/or paresis of hindlimbs; 3 = paralysis of hindlimbs and/or paresis of forelimbs; 4 = tetraparesis; 5 = moribund or death) were recorded daily. Cells were injected by means of the cisterna magna (i.c.) as previously described¹⁸, whereas i.v. cell injections were performed through the tail vein using a 25-gauge needle. All procedures involving animals were performed according to the guidelines of the Animal Ethical Committee of our Institute.

Histology

Brain, spinal cord and peripheral organs (heart, lung, liver, spleen, gut, kidney) from mice transcardially perfused with PBS followed by 4% paraformaldehyde were removed and processed for light and electron microscopy. Paraffin-embedded tissue sections (5 μ m) were stained with haematoxylin and eosin, Luxol fast blue and Bielschowsky to detect inflammatory infiltrates, demyelination and axonal loss, respectively¹⁸. Quantification of CNS damage was performed using IM-50 image analyser software (Leica) on 300 sections.

To detect the *in vivo* fate of neural precursors in i.v.- and i.c.-injected mice, frozen (10–12 μ m) and fresh agarose-embedded tissue sections (50–80 μ m) were cut and incubated overnight at 37 °C in 5-bromo-4-chloro-3-indolyl- β -D-galactoside (X-gal) solution for nuclear β -gal activity, and processed for immunohistochemistry or counterstained with nuclear red. Sections of CNS showing β -gal⁺ cells were re-cut at 5 μ m and double-stained using anti-glia fibrillary acidic protein (GFAP) for astrocytes (Dako), anti-neuronal nuclear antigen (NeuN) for neurons (Chemicon), anti-PDGFR α receptor for oligodendrocyte progenitors (SantaCruz) and anti-*nestin* (Chemicon) for undifferentiated cells. Appropriate biotin-conjugated (Amersham) secondary antibodies were used. A total of 7,274 double-positive cells were counted on 480 5- μ m sections of CNS (31 brain and 17 spinal cord sections per mouse; five mice per group) taken at 100- μ m intervals. Light and electron microscopy was performed on semi-thin serial sections (1 μ m) from tissues stained with 5-bromo-3-indolyl-beta-D-galactopyranoside, post-fixed in 2% glutaraldehyde and 1% osmium tetroxide, and embedded in Epon as previously described (see also Supplementary Information)³². Confocal microscopy (Bio Rad, MRC 1024) was performed on fresh-frozen brain and spinal cord tissue sections (5 μ m) obtained from EAE mice injected either i.v. (three mice) or i.c. (three mice) with 1×10^6 of the same neural precursors, transfected with the same lentiviral vector as above but containing the enhanced green fluorescent protein (eGFP) gene in place of the *nls-lacZ* gene³¹. Tissue sections were analysed at 0.5- μ m intervals using the same primary antibodies as above and rhodamine-conjugated secondary antibodies.

Semi-quantitative RT-PCR

Total RNA was extracted from the brain and spinal cord of a minimum of five mice per group and from neural precursors undifferentiated and *in vitro* differentiated for 10 days after growth factor removal, using RNAfast (Molecular System), and complementary

DNA was synthesized. Polymerase chain reaction with reverse transcription (RT-PCR) was performed and PCR products were hybridized with fluorescein-labelled probes. Enhanced chemical fluorescence signal-amplification module (Amersham Pharmacia Biotech) was used; signals were detected using a high-performance laser scanning system (Typhoon 8600, Amersham Pharmacia Biotech). Values were normalized against the *Gapdh* gene and expressed as arbitrary units (AU). Primer and probes have been designed using Primer Express 1.5 software (Applied Biosystem) (see also Supplementary Information).

FACS analysis

Neural precursors were washed and then incubated with fluorescein-isothiocyanate-conjugated anti-mouse CD44, L-selectin, LFA-1, PSGL-1 and VLA-4 antibodies. Isotype control was performed using fluorescein isothiocyanate-labelled mouse control immunoglobulin- γ . Analysis was performed with a FACScan flow cytometer (Becton Dickinson) equipped with CellQuest software, and 50,000 events were acquired.

Neurophysiological assessment

Mice were anaesthetized with tribromoethanol (0.02 ml g^{-1} of body weight), and placed under a heating lamp to avoid hypothermia. Motor-evoked potentials (MEPs)—muscle responses evoked by transcranial electrical stimulation of the motor cortex—were recorded using a pair of 25-gauge needle electrodes from the contralateral lower limb. The active electrode was inserted into the footpad muscles, whereas the reference was placed subcutaneously between the first and the second digit. The excitatory volleys descending along the cortico-spinal pathways evoke motor potentials in the paw muscles through a trans-synaptic depolarization of alpha motor neurons (cortical MEP). A spinal MEP was also obtained stimulating ventral nerve roots over the lumbar spine, so that a CCT was measured as time difference—expressed in milliseconds (ms)—between cortical and spinal MEP latencies.

Statistical analysis

Data were compared using the Student's *t*-test for unpaired data, the Mann-Whitney *U*-test for non-parametric data, or the χ^2 test.

Received 13 November 2002; accepted 10 March 2003; doi:10.1038/nature01552.

- Lucchinetti, C. *et al.* Heterogeneity of multiple sclerosis lesions: implications for the pathogenesis of demyelination. *Ann. Neurol.* **47**, 707–717 (2000).
- Hemmer, B., Archelos, J. J. & Hartung, H. P. New concepts in the immunopathogenesis of multiple sclerosis. *Nature Rev. Neurosci.* **3**, 291–301 (2002).
- Archer, D. R., Cuddon, P. A., Lipsitz, D. & Duncan, I. D. Myelination of the canine central nervous system by glial cell transplantation: a model for repair of human myelin disease. *Nature Med.* **3**, 54–59 (1997).
- Groves, A. K. *et al.* Repair of demyelinated lesions by transplantation of purified O-2A progenitor cells. *Nature* **362**, 453–455 (1993).
- Blakemore, W. F. Remyelination of CNS axons by Schwann cells transplanted from the sciatic nerve. *Nature* **266**, 68–69 (1977).
- Imaizumi, T., Lankford, K. L., Burton, W. V., Fodor, W. L. & Kocsis, J. D. Xenotransplantation of transgenic pig olfactory ensheathing cells promotes axonal regeneration in rat spinal cord. *Nature Biotechnol.* **18**, 949–953 (2000).
- Jefferson, S. *et al.* Inhibition of oligodendrocyte precursor motility by oligodendrocyte processes: implications for transplantation-based approaches to multiple sclerosis. *Mult. Scler.* **3**, 162–167 (1997).
- Franklin, R. J. & Blakemore, W. F. To what extent is oligodendrocyte progenitor migration a limiting factor in the remyelination of multiple sclerosis lesions? *Mult. Scler.* **3**, 84–87 (1997).
- Clarke, D. & Frisen, J. Differentiation potential of adult stem cells. *Curr. Opin. Genet. Dev.* **11**, 575–580 (2001).
- Horner, P. J. & Gage, F. H. Regenerating the damaged central nervous system. *Nature* **407**, 963–970 (2000).
- Teng, Y. D. *et al.* Functional recovery following traumatic spinal cord injury mediated by a unique polymer scaffold seeded with neural stem cells. *Proc. Natl Acad. Sci. USA* **99**, 3024–3029 (2002).
- Gritti, A. *et al.* Epidermal and fibroblast growth factors behave as mitogenic regulators for a single multipotent stem cell-like population from the subventricular region of the adult mouse forebrain. *J. Neurosci.* **19**, 3287–3297 (1999).
- Galli, R. *et al.* Emx2 regulates the proliferation of stem cells of the adult mammalian central nervous system. *Development* **129**, 1633–1644 (2002).
- Deckert-Schluter, M., Schluter, D., Hof, H., Wiestler, O. D. & Lassmann, H. Differential expression of ICAM-1, VCAM-1 and their ligands LFA-1, Mac-1, CD43, VLA-4, and MHC class II antigens in murine *Toxoplasma encephalitis*: a light microscopic and ultrastructural immunohistochemical study. *J. Neuropathol. Exp. Neurol.* **53**, 457–468 (1994).
- Butcher, E. C. & Picker, L. J. Lymphocyte homing and homeostasis. *Science* **272**, 60–66 (1996).
- Brocke, S., Piercy, C., Steinman, L., Weissman, I. L. & Veromaa, T. Antibodies to CD44 and integrin $\alpha 4$, but not L-selectin, prevent central nervous system inflammation and experimental encephalomyelitis by blocking secondary leukocyte recruitment. *Proc. Natl Acad. Sci. USA* **96**, 6896–6901 (1999).
- Del Maschio, A. *et al.* Leukocyte recruitment in the cerebrospinal fluid of mice with experimental meningitis is inhibited by an antibody to junctional adhesion molecule (JAM). *J. Exp. Med.* **190**, 1351–1356 (1999).
- Furlan, R. *et al.* Intrathecal delivery of IFN- γ protects C57BL/6 mice from chronic-progressive experimental autoimmune encephalomyelitis by increasing apoptosis of central nervous system-infiltrating lymphocytes. *J. Immunol.* **167**, 1821–1829 (2001).
- Gensert, J. M. & Goldman, J. E. Endogenous progenitors remyelinate demyelinated axons in the adult CNS. *Neuron* **19**, 197–203 (1997).
- Elanicher, F., Kehrli, P., Labourette, G. & Sensenbrenner, M. Basic fibroblast growth factor (bFGF) injection activates the glial reaction in the injured adult rat brain. *Brain Res.* **737**, 201–214 (1996).
- Moon, L. D. & Fawcett, J. W. Reduction in CNS scar formation without concomitant increase in axon regeneration following treatment of adult rat brain with a combination of antibodies to TGF β 1 and β 2. *Eur. J. Neurosci.* **14**, 1667–1677 (2001).
- Linker, R. A. *et al.* CNTF is a major protective factor in demyelinating CNS disease: a neurotrophic cytokine as modulator in neuroinflammation. *Nature Med.* **8**, 620–624 (2002).
- Alexander, J. S. & Elrod, J. W. Extracellular matrix, junctional integrity and matrix metalloproteinase interactions in endothelial permeability regulation. *J. Anat.* **200**, 561–574 (2002).
- Temple, S. The development of neural stem cells. *Science* **414**, 112–117 (2001).
- Franklin, R. J. Why does remyelination fail in multiple sclerosis? *Nature Rev. Neurosci.* **3**, 705–714 (2002).
- Kim, J. H. *et al.* Dopamine neurons derived from embryonic stem cells function in an animal model of Parkinson's disease. *Nature* **418**, 50–56 (2002).
- Aboody, K. S. *et al.* Neural stem cells display extensive tropism for pathology in adult brain: evidence from intracranial gliomas. *Proc. Natl Acad. Sci. USA* **97**, 12846–12851 (2000).
- Choop, M. & Li, Y. Treatment of neural injury with marrow stromal cells. *Lancet Neurol.* **1**, 92–100 (2002).
- Learish, R. D., Brustle, O., Zhang, S. C. & Duncan, I. D. Intraventricular transplantation of oligodendrocyte progenitors into a fetal myelin mutant results in widespread formation of myelin. *Ann. Neurol.* **46**, 716–722 (1999).
- Chen, J. *et al.* Therapeutic benefit of intracerebral transplantation of bone marrow stromal cells after cerebral ischemia in rats. *J. Neurol. Sci.* **189**, 49–57 (2001).
- Follenzi, A., Ailles, L. E., Bakovic, S., Geuna, M. & Naldini, L. Gene transfer by lentiviral vectors is limited by nuclear translocation and rescued by HIV-1 pol sequences. *Nature Genet.* **25**, 217–222 (2000).
- Feltri, M. L. *et al.* Conditional disruption of beta1 integrin in Schwann cells impedes interactions with axons. *J. Cell. Biol.* **156**, 199–209 (2002).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements We thank L. De Filippis and L. Naldini for providing the lentiviral vectors and G. Constantin and B. Rossi for contributing to FACS analysis. We also thank C. Panzeri for technical help with confocal microscopy. This work was supported by the Italian Multiple Sclerosis Association (AISM), Myelin Project, European Union (EU), Fondazione Agnini and BMW Italia.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to G.M. (e-mail: martino.gianvito@hsr.it) or A.L.V. (e-mail: vescovi@tin.it).

A median redshift of 2.4 for galaxies bright at submillimetre wavelengths

S. C. Chapman*, A. W. Blain*, R. J. Ivison† & Ian R. Smail‡

* California Institute of Technology, Pasadena, California 91125, USA

† Astronomy Technology Centre, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK

‡ Institute for Computational Cosmology, University of Durham, South Road, Durham DH1 3LE, UK

A significant fraction of the energy emitted in the early Universe came from very luminous galaxies that are largely hidden at optical wavelengths (because of interstellar dust grains); this energy now forms part of the cosmic background radiation at wavelengths near 1 mm (ref. 1). Some submillimetre (submm) galaxies have been resolved from the background radiation², but they have been difficult to study because of instrumental limitations³. This has impeded the determination of their redshifts (z), which is a crucial element in understanding their nature and evolution⁴. Here we report spectroscopic redshifts for ten submm galaxies that were identified using high-resolution radio observations^{5–7}. The median redshift for our sample is 2.4, with a quartile range of 1.9–2.8. This population therefore coexists with the peak activity of quasars, suggesting a close relationship between the growth of massive black holes and luminous dusty galaxies⁸. The space density of submm galaxies at redshifts over 2 is about 1,000 times greater than that of similarly luminous galaxies in the present-day Universe, so they represent an important component of star formation at high redshifts.

The first real insight in the origin of the far-infrared/submm background came with the commissioning of the Submillimeter Common User Bolometric Array (SCUBA) camera on the James Clerk Maxwell Telescope (JCMT). Almost 100% of the submm background has been resolved with the deepest SCUBA maps, which exploit a sensitivity boost from gravitational lenses^{2,3,9}. However, the near-infrared/optical faintness of the submm galaxy population conspires with the modest positional precision available from the large (15") SCUBA beam and the large surface density of unrelated, optically faint galaxies to render positional coincidence alone inadequate to identify counterparts^{3,5}. This problem is compounded by the small field of view of SCUBA: compiling large samples of submm galaxies is slow, with roughly one detected every night. Because of these difficulties, robust spectroscopic redshifts have been published for only a handful of submm sources, all atypically bright at optical wavelengths^{3,10–12}.

Here, we have overcome these problems by taking advantage of 1.4-GHz radio data from the Very Large Array (VLA) radio telescope (Hubble Deep Field¹³, SSA13 field¹⁴, ElaisN2 and Lockman fields⁷). These deep images have fine (1–2") spatial resolution and a large (30' full-width at half-maximum, FWHM) field of view. The radio data was exploited to follow up SCUBA submm data in the Hubble Deep Field (12-h RA) and SSA13 (13-h RA) fields^{5,15,16}, and the Lockman (10-h RA) and ElaisN2 (16-h RA) fields¹⁷. The radio data are sensitive to the synchrotron radio emission from cosmic-ray electrons accelerated in the supernovae explosions of the same high-mass stars that heat the dust and generate the far-infrared/submm emission. Hence using a deep radio image should be an efficient way of pinpointing the positions of many submm galaxies. This is confirmed in fields with both submm and radio observations: radio counterparts brighter than 30 μ Jy can be found for 60–70% of submm galaxies brighter than 5 mJy at 850 μ m (which have a surface density³ of 450 deg⁻², and contribute about 20% of the submm background)^{7,16}. Moreover, the accurate radio positions mean that SCUBA's efficient 'photometry' mode can be used to search for submm galaxies^{5,6,18}, raising the detection rate to around ten per night.

Although submm galaxies have faint optical continuum emission³, the radio-detected samples are sufficiently large to allow efficient, multi-object optical spectroscopy using 10-m telescopes. Slit positions can be assigned accurately, as the radio/submm emission can be located on optical images to within 0.3–0.8" (ref. 7). Using this approach we targeted 34 radio-detected submm galaxies, mostly brighter than 5 mJy at 850 μ m with optical magnitudes in the range $I = 22.2$ – 26.4 , using the Echelle Spectrograph and Imager (ESI) and the Low-Resolution and Imaging Spectrograph (LRIS) on the Keck telescopes (Fig. 1). We measured redshifts for ten submm galaxies, representative of the blank-field population. We have doubled the number of optically bright ($I < 23.5$) submm galaxies with accurate redshifts. Our six redshifts for $I > 23.5$ submm galaxies represent the first redshifts for this class of submm galaxy. The redshifts span the range $z = 0.8$ – 3.7 , with a median of 2.4 and an interquartile range of 1.9–2.8 (Fig. 2). These redshifts allow dust temperatures T_d and bolometric luminosities to be determined (see Table 1), assuming that the tight correlation between far-infrared and radio emission observed at low redshifts¹⁹ remains valid. Typical values are of order 35 K and several $10^{12} L_\odot$ respectively (where $L_\odot$ is the luminosity of the Sun). Note that relativistic electrons accelerated in shocks close to an active galactic nucleus (AGN) should boost the radio flux density above that expected from the standard correlation. If the radio emission is boosted by an AGN then both temperature and luminosity are overestimated. There are no obvious cases in the present sample in this category.

Table 1 Positions and parameters of the observed submillimetre galaxies

Name	Redshift	$S_{850 \mu m}$ (mJy)	I (mag)	T_d (K)	L_{bol} ($10^{12} L_\odot$)
SMM105224.6+572119	2.429	11.7 $\pm$ 3.4*	25.9	21 $\pm$ 5	3.0 $^{+2.0}_{-0.7}$
SMM163704.3+410530	0.840	11.2 $\pm$ 1.6*	21.5	16 $^{+4.1}_{-2.2}$	0.58 $^{+0.32}_{-0.34}$
SMM105230.6+572212	2.610	11.0 $\pm$ 2.6*	23.0	29 $\pm$ 6	6.9 $^{+5.9}_{-3.3}$
SMM163650.0+405733	2.376	8.2 $\pm$ 1.7*	22.2	54 $^{+11}_{-7}$	45 $^{+53}_{-19}$
SMM123600.2+621047	1.998	7.9 $\pm$ 2.4	23.6	37 $^{+8}_{-4}$	10 $^{+11}_{-4}$
SMM131201.2+424208	3.419	6.2 $\pm$ 1.2	23.5	48 $^{+10}_{-7}$	20 $^{+18}_{-7}$
SMM105207.7+571907	2.698	6.2 $\pm$ 1.6*	26.0	36 $^{+7}_{-4}$	7.8 $^{+7.1}_{-2.3}$
SMM131212.7+424423	2.811	5.6 $\pm$ 1.9	26.4	54 $^{+19}_{-10}$	30 $^{+60}_{-16}$
SMM163658.8+405733	1.189	5.1 $\pm$ 1.4*	22.4	25 $^{+5}_{-9}$	1.5 $\pm$ 1.2
SMM105155.7+572312	3.699	4.5 $\pm$ 1.3*	24.2	59 $^{+26}_{-16}$	30 $^{+82}_{-20}$

Positions of submm galaxies with redshifts from our survey, along with their measured redshifts, submm fluxes ($S_{850 \mu m}$) and optical magnitudes (I), inferred dust temperatures (T_d) and bolometric luminosities (L_{bol}). T_d and L_{bol} were calculated from the redshift, submm and radio flux densities, assuming that the far-infrared-radio correlation applies. Sources are in order of decreasing flux density, as in Fig. 1. (SMM163704.3+410530 was previously published in ref. 26.) Submm fluxes for sources in the Lockman (10-h RA) and ElaisN2 (16-h RA) fields (indicated by an asterisk) are from ref. 17 and we measured the fluxes for other sources. Six out of ten galaxies have I -band magnitudes fainter than 23.5, consistent with the optical flux distribution of blank-field submm galaxies presented elsewhere¹⁶, and confirming that our sample is representative of the population. The relationship between submm galaxies and other high-redshift populations will be better understood by comparing the strength of the clustering, both within and between the different classes. This will require substantially larger samples than are currently available.

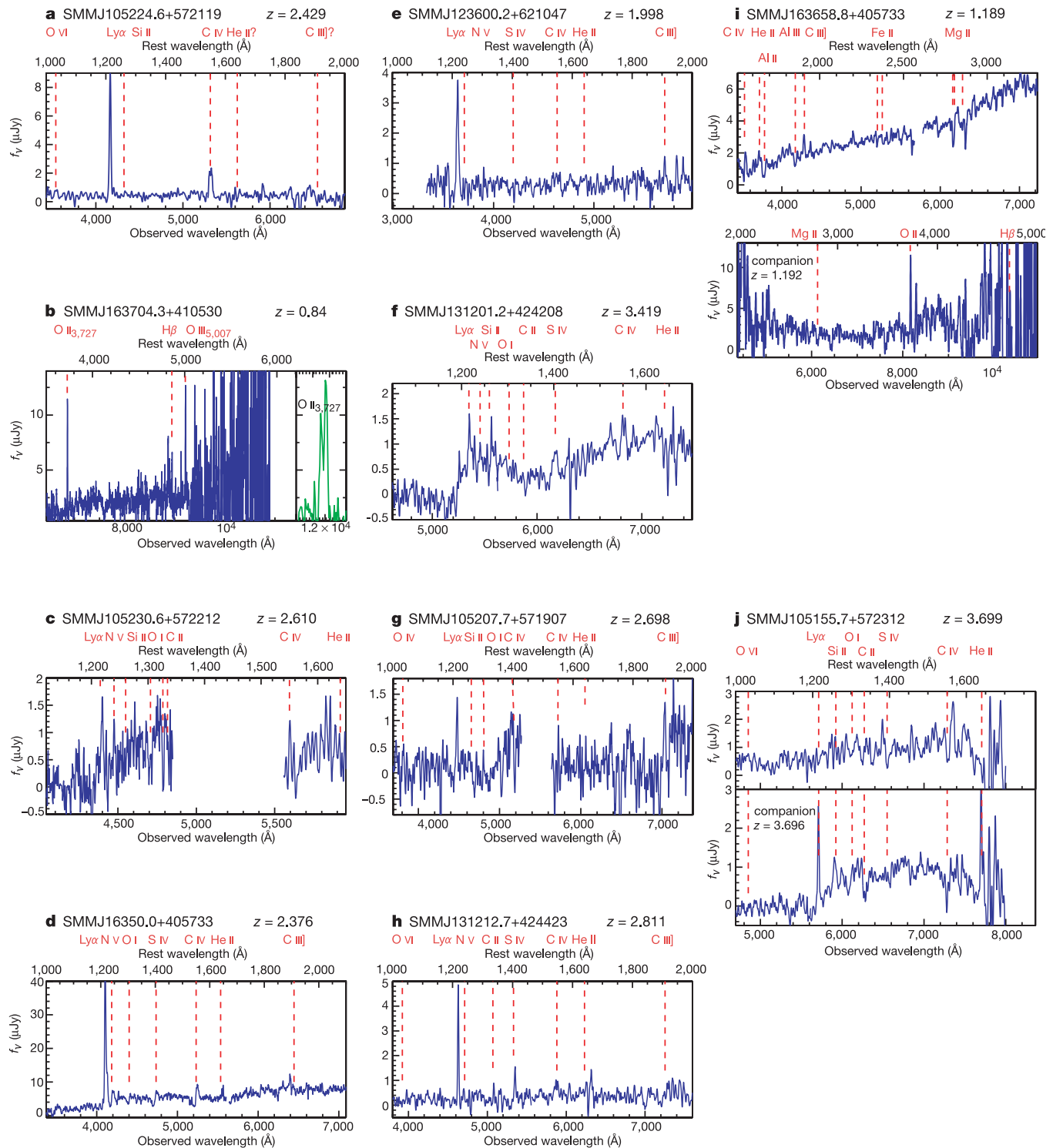


Figure 1 The ten spectra (a–j) from Table 1 with identified redshifts from the sample of 34 submm galaxies observed, spanning the $l = 22.2$ – 26.4 range. The redshifts are smoothed to the instrumental resolution. Pilot observations, using Keck's ESI²⁷ spectrograph on 16 July 2001 in the ELAIS-N2 field¹⁷ yielded three spectra with strong emission lines. Multi-object Keck observations of 31 submm galaxies using LRIS²⁸ were made on 18–19 March 2002 in light cirrus conditions (with good $0.8''$ seeing) in the HDF and SSA13 fields^{5,16}, and the Lockman Hole and ELAIS-N2 fields¹⁷. All the spectra probe the observed wavelength range from 0.3 – $1 \mu\text{m}$. Exposure times were 1.5 – 4.5 h, split into 30-min integrations. Data reduction followed standard techniques using custom Image Reduction Astronomy Facility (IRAF) scripts. One-dimensional spectra were extracted and compared with template spectra and emission-line catalogues to identify redshifts. All identifications are based on multiple lines, most prominently the Ly α line which varies tremendously in both flux (ranging from 1 to $60 \mu\text{Jy}$) and rest-frame equivalent width (3 to

$>100 \text{\AA}$). Weaker stellar/interstellar/AGN features and/or continuum breaks were detected in the spectra, strengthening the redshift identifications. Two galaxies at the same redshifts (for sources i, j) were identified in the vicinity of the ten robust submm sources, as shown for the companions SMMJ105155.7 + 572312-a (one of the few submm galaxies with redshifts showing Lyman- α in absorption) and SMMJ105155.7 + 572312-b. **d**, The highest signal-to-noise spectrum (SMMJ163650 + 405733; ref. 30) includes features indicative of both strong starburst activity (P Cygni wind absorption profiles, stellar/interstellar absorption lines) and a weak AGN, a mix typical of hybrid starburst/Seyfert-2 galaxies. None of our spectra have broad lines like Type-I AGN. Narrow-line Type-II AGN with enhanced N v and/or C iv emission are consistent with half of our spectra (a, d, f, g, h). Despite this evidence for AGN in these galaxies, AGN appear to be relatively weak in energetic terms as compared to those identified in X-ray or optical surveys. f_v , specific flux density.

The redshifts impose robust limits to the space densities of submm galaxies in the redshift ranges 0.5–1.2, 1.8–2.8 and 2.8–4 (the gap at $z = 1.2$ –1.8 is the ‘spectroscopic desert’ in which no strong restframe ultraviolet lines are redshifted into the optical window): we calculate $\rho = (3.3 \pm 2.3) \times 10^{-6} \text{ Mpc}^{-3}$, $(6.5 \pm 2.5) \times 10^{-6} \text{ Mpc}^{-3}$ and $(2.4 \pm 1.2) \times 10^{-6} \text{ Mpc}^{-3}$, respectively, assuming a flat (cosmological constant $\Lambda = 0.7$) Universe with current Hubble constant $H_0 = 65 \text{ km s}^{-1} \text{ Mpc}^{-1}$. This compares to a space density of around 10^{-8} Mpc^{-3} for low-redshift galaxies with comparable bolometric luminosities greater than $5 \times 10^{12} L_\odot$, comparable to the luminosities of our spectroscopically targeted galaxies at any redshift greater than one. The volume density of very luminous galaxies thus increases almost a thousandfold from $z \approx 0$ to 2.

The relationship between different classes of high-redshift galaxy is critical for our understanding of galaxy evolution. Our spectroscopy shows that the submm galaxies are coeval with the important populations of star-forming galaxies and quasars detected at $z = 2$ –3 using optical and X-ray techniques^{8,20–22}. We can use their relative space densities to compare and relate these populations. In a $1,000 \text{ Mpc}^3$ box at $z \approx 2.5$, there are ten Lyman-break galaxies with $R_{\text{AB}} < 25.5$ (about $0.2 \times L_{1,500}^*$) and one submm galaxy from our sample (with $850\text{-}\mu\text{m}$ flux $> 5 \text{ mJy}$), but the single submm galaxy produces a comparable luminosity to the ten Lyman-break galaxies. (Here R_{AB} is the flux normalized magnitude near $6,900 \text{ \AA}$; $L_{1,500}^*$ is the characteristic luminosity at restframe $1,500 \text{ \AA}$, denoting a change in slope in the luminosity function.) Equally, the volume density of submm galaxies is similar to that of narrow-line Type-II AGN selected at these redshifts, from both X-ray and optical surveys, and is an order of magnitude greater than that of bright optical quasars with bolometric magnitude $M_B > -25$ (ref. 8). However, the links between the different classes

of sources all depend critically on whether the activity is intermittent or continuous.

To quantify the whole submm population using this survey we must account for several selection effects. In particular, requiring a radio identification prior to spectroscopy limits the maximum redshift (Fig. 2). Studies of galaxies at low and moderate redshifts suggest that changes in the dust properties, especially the dust temperature T_d , modify the relative strength of the emission in the submm and radio wavebands²³. The detectability of distant submm galaxies (at a fixed $850\text{-}\mu\text{m}$ flux) in the radio waveband thus depends on the dust temperature: warmer galaxies are detectable at $z \geq 3$, while the radio emission from cooler sources falls below the flux limit at lower redshifts¹⁸. About 30% of submm galaxies that are brighter than 5 mJy at $850\text{-}\mu\text{m}$ are not detected at radio wavelengths: these galaxies would be lost from our catalogue at $z \approx 2.5$ for the coldest plausible $T_d \approx 25 \text{ K}$, and at $z > 3$ for the typical $T_d \approx 35 \text{ K}$ that we find. Hence, we can place only a lower limit on the space density in the highest redshift interval. The density derived at $z = 2$ should be unaffected by the galaxies’ detection at radio wavelengths.

We must also consider the completeness of our spectroscopic observations. Although our redshift completeness is relatively high ($\sim 30\%$), we failed to find convincing redshifts for 24 galaxies. However, nine of these have strong single emission line detections which could plausibly be identified as Lyman- α . Some of the remainder without emission lines could lie in the spectroscopic desert, making redshifts difficult to identify. The absence of sources at $z \approx 1.5$ in Table 1 suggests that this effect is probably responsible for some of the incompleteness. Our targets generally have faint ultraviolet continua, and so there is a bias towards identifying redshifts for galaxies with strong emission lines. If the missing galaxies lie outside the spectroscopic desert, but lack strong emission lines, then their radio detection suggests that they probably lie in the same $z = 1$ –4 range spanned by the spectroscopically identified sources.

On the basis of bolometric luminosities estimated from our radio and submm measurements and redshifts (Table 1), assuming the far-infrared–radio correlation, we calculate the following luminosity densities from our survey: $\rho_L(0.5 < z < 1.2) = (5 \pm 5) \times 10^6 L_\odot \text{ Mpc}^{-3}$, $\rho_L(1.8 < z < 2.8) = (7^{+6}_{-4}) \times 10^7 L_\odot \text{ Mpc}^{-3}$ and $\rho_L(2.8 < z < 4.0) = (7^{+10}_{-5}) \times 10^7 L_\odot \text{ Mpc}^{-3}$. To translate these values into star-formation densities we must subtract any contribution to the luminosities of these galaxies from the heating of dust by an active galactic nucleus (AGN). The proportion of the submm population detected in deep Chandra and XMM-Newton X-ray surveys, which are sensitive to even dust-obscured AGN^{7,20}, at flux densities significantly greater than that expected from star formation alone is at most 30%. Hence, we assume conservatively that the far-infrared luminosity is derived from star formation in the remaining 70%. Luminosity-density measurements can be translated into a star-formation density using the standard calibration²⁴ of 1.6 – $2.2 \times 10^9 L_\odot (M_\odot \text{ yr}^{-1})^{-1}$.

The bright radio-detected submm galaxies presented here represent just 20% of the $850\text{-}\mu\text{m}$ background, yet their estimated star-formation densities at redshifts of 2 and 3 are comparable with those inferred from optical observations²¹. Both of these estimates require corrections: for the optical surveys an increase of a factor of 5 is needed to account for dust obscuration²⁵, and our submm estimate needs to be corrected by a similar factor to account for spectroscopic incompleteness and sources below our submm flux limit. We find that the submm and Lyman-break populations encompass all of the star formation activity needed to reproduce the star formation history inferred from models of the background radiation and submm counts⁴. Hence, our substantially complete redshift survey provides the strongest evidence for a rapid increase in the density of dusty, very luminous galaxies in the distant Universe. Their density peaks at a redshift $z \approx 2.4$ at a value about 1,000 times greater than

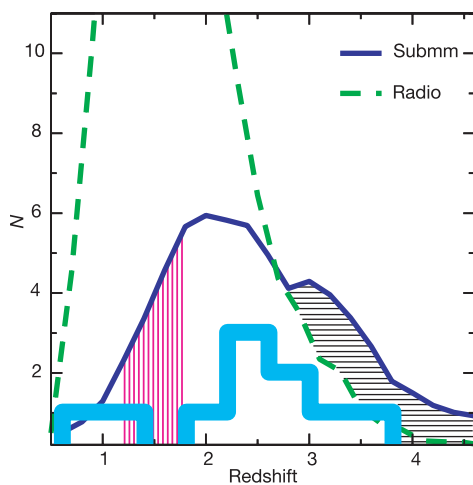


Figure 2 The redshift histogram of our submm galaxy sample (heavy cyan line). We describe the selection effects using an evolving model of the local far-infrared luminosity function²⁹, in which the dusty galaxies are represented by a range of template spectral energy distributions. The model has been tuned to fit the statistical properties of the submm-radio galaxy population (surface density on the sky and submm–radio colours²⁹). Because a range of spectral energy distributions are represented at each galaxy luminosity, a broader redshift distribution results than if the luminosity were tied one-to-one with the dust temperature²³. We plot the predicted redshift distributions for submm galaxies with flux densities $S_{850\text{-}\mu\text{m}} > 5 \text{ mJy}$ (solid blue line) and radio sources with $30\text{-}\mu\text{m} < S_{1.4\text{ GHz}} < 500\text{-}\mu\text{m}$ (dashed green line). We expect to miss sources lying between the submm and radio model curves (shaded black) owing to our requirement of a radio detection to pinpoint the submm source. The apparent deficit (shaded magenta) between model and data at $z \approx 1.5$ may be indicative of the difficulty in obtaining spectra in the spectroscopic desert that spans the range $1.2 < z < 1.8$. Note that the models are in good overall agreement with the observed redshift distribution.

the density at the present epoch. Thus we conclude that the balance between the star-formation density in the most luminous dust-obscured versus unobscured galaxies has altered radically over the last 80% of the history of the Universe. □

Received 1 July 2002; accepted 17 February 2003; doi:10.1038/nature01540.

- Fixsen, D. J., Dwek, E., Mather, J. C., Bennett, C. L. & Shafer, R. A. The spectrum of the extragalactic far-infrared background from the COBE FIRAS observations. *Astrophys. J.* **508**, 123–128 (1998).
- Smail, I., Ivison, R. J. & Blain, A. W. A deep submillimeter survey of lensing clusters: a new window on galaxy formation and evolution. *Astrophys. J.* **490**, L5–L8 (1997).
- Smail, I., Ivison, R. J., Blain, A. W. & Kneib, J.-P. The nature of faint submillimetre-selected galaxies. *Mon. Not. R. Astron. Soc.* **331**, 495–520 (2002).
- Blain, A. W., Smail, I., Ivison, R. J. & Kneib, J.-P. The history of star-formation in dusty galaxies. *Mon. Not. R. Astron. Soc.* **302**, 632–648 (1999).
- Barger, A. J., Cowie, L. L. & Richards, E. A. Mapping the evolution of high-redshift dusty galaxies with submillimeter observations of a radio-selected sample. *Astron. J.* **119**, 2092–2109 (2000).
- Chapman, S. C., Richards, E. A., Lewis, G. F., Wilson, G. & Barger, A. J. The nature of the bright submillimeter galaxy population. *Astrophys. J.* **548**, L147–L151 (2001).
- Ivison, R. J. *et al.* Deep radio imaging of the SCUBA 8-mJy survey fields: submm source identifications and redshift distribution. *Mon. Not. R. Astron. Soc.* **337**, 1–25 (2002).
- Boyle, B. J. *et al.* The 2dF QSO redshift survey—I. The optical luminosity function of quasi-stellar objects. *Mon. Not. R. Astron. Soc.* **317**, 1014–1022 (2000).
- Cowie, L. L., Barger, A. J. & Kneib, J.-P. Faint submillimeter counts from deep 850 micron observations of the lensing clusters A370, A851, and A2390. *Astron. J.* **123**, 2197–2205 (2002).
- Ivison, R. J. *et al.* A hyperluminous galaxy at $z = 2.8$ found in a deep submillimetre survey. *Mon. Not. R. Astron. Soc.* **298**, 583–593 (1998).
- Barger, A. J. *et al.* Redshift distribution of the faint submillimeter galaxy population. *Astron. J.* **117**, 2656–2665 (1999).
- Ivison, R. J. *et al.* The diversity of SCUBA-selected galaxies. *Mon. Not. R. Astron. Soc.* **315**, 209–221 (2000).
- Richards, E. A. The nature of radio emission from distinct galaxies: The 1.4 GHz observations. *Astrophys. J.* **533**, 611–630 (2000).
- Fomalont, E. B., Kellermann, K. I., Partridge, R. B., Windhorst, R. A. & Richards, E. A. The microjansky sky at 8.4 GHz. *Astron. J.* **123**, 2402–2416 (2002).
- Barger, A. J. *et al.* Submillimetre-wavelength detection of dusty star-forming galaxies at high redshift. *Nature* **394**, 248–251 (1998).
- Chapman, S. C. *et al.* The properties of microjansky radio sources in the HDF-N, SSA13, and SSA22 fields. *Astrophys. J.* **585**, 57–66 (2003).
- Scott, S. E. *et al.* The SCUBA 8-mJy survey—I. Submillimetre maps, sources and number counts. *Mon. Not. R. Astron. Soc.* **331**, 817–837 (2002).
- Chapman, S. C., Lewis, G. F., Scott, D., Borys, C. & Richards, E. A. Understanding the nature of the optically faint radio sources and their connection to the submillimeter population. *Astrophys. J.* **570**, 557–570 (2002).
- Condon, J. Radio emission from ordinary galaxies. *Annu. Rev. Astron. Astrophys.* **30**, 575–611 (1992).
- Barger, A. J. *et al.* X-ray, optical and infrared imaging and spectral properties of the 1-Ms Chandra Deep Field North sources. *Astron. J.* **124**, 1839–1885 (2002).
- Steidel, C. C., Adelberger, K. L., Gialavalis, M., Dickinson, M. & Pettini, M. Lyman-break galaxies at $z > 4$ and the evolution of the ultraviolet luminosity density at high redshift. *Astrophys. J.* **519**, 1–17 (1999).
- Steidel, C. C. *et al.* The population of faint optically-selected AGN at $z \sim 3$. *Astrophys. J.* (in the press); preprint available at <http://www.arXiv.org/astro-ph/0205142> (2002).
- Blain, A. W., Smail, I., Ivison, R. J., Kneib, J.-P. & Frayer, D. T. Submillimeter galaxies. *Phys. Rep.* **369**, 111–196 (2002); preprint available at <http://www.arXiv.org/astro-ph/0202228> (2002).
- Kennicutt, R. C. Star formation in galaxies along the Hubble sequence. *Annu. Rev. Astron. Astrophys.* **36**, 189–232 (1998).
- Pettini, M. *et al.* The rest-frame optical spectra of Lyman-break galaxies: star formation, extinction, abundances and kinematics. *Astrophys. J.* **554**, 981–1000 (2001).
- Chapman, S. C., Smail, I., Ivison, R. & Blain, A. The effect of lensing on the identification of SCUBA galaxies. *Mon. Not. R. Astron. Soc.* **335**, 17–21 (2002).
- Sheinis, A. I. *et al.* ESI, a new Keck Observatory echellette spectrograph and imager. *Publ. Astron. Soc. Pacif.* **114**, 851–865 (2002).
- Oke, J. B. *et al.* The Keck low-resolution imaging spectrometer. *Publ. Astron. Soc. Pacif.* **107**, 375–385 (1995).
- Chapman, S. C., Helou, G., Lewis, G. F. & Dale, D. The bi-variate luminosity-colour distribution of IRAS galaxies, and implications for the high redshift Universe. *Astrophys. J.* (in the press); preprint available at <http://www.arXiv.org/astro-ph/0301233> (2002).
- Smail, I. *et al.* A vigorous starburst in the SCUBA galaxy N2_850.4. *Mon. Not. R. Astron. Soc.* (in the press); preprint available at <http://www.arXiv.org/astro-ph/0303128> (2003).

Acknowledgements We thank C. Steidel, A. Shapley and T. Heckman for discussions. S.C.C. acknowledges support from NASA. I.R.S. acknowledges support from the Royal Society and a Philip Leverhulme Prize Fellowship. NRAO is operated by Associated Universities Inc., under a cooperative agreement with the US National Science Foundation. Data presented herein were obtained using the W. M. Keck Observatory, which is operated as a scientific partnership among Caltech, the University of California and NASA. The Observatory was made possible by the financial support of the W. M. Keck Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.C.C. (e-mail: schapman@irastro.caltech.edu).

A hidden pseudogap under the ‘dome’ of superconductivity in electron-doped high-temperature superconductors

L. Alff*, Y. Krockenberger*, B. Welter*, M. Schonecke*, R. Gross*, D. Manske† & M. Naito‡

* Walther-Meissner-Institut, Bayerische Akademie der Wissenschaften, 85748 Garching, Germany

† Freie Universität Berlin, Institut für Theoretische Physik, 14195 Berlin, Germany

‡ NTT Basic Research Laboratories, 3-1 Morinosato, Atsugi, Kanagawa 243, Japan

The ground state of superconductors is characterized by the long-range order of condensed Cooper pairs: this is the only order present in conventional superconductors. The high-transition-temperature (high- T_c) superconductors, in contrast, exhibit more complex phase behaviour, which might indicate the presence of other competing ground states. For example, the pseudogap^{1,2}—a suppression of the accessible electronic states at the Fermi level in the normal state of high- T_c superconductors—has been interpreted as either a precursor to superconductivity^{3,4} or as tracer of a nearby ground state that can be separated from the superconducting state by a quantum critical point^{5,6}. Here we report the existence of a second order parameter⁷ hidden within the superconducting phase of the underdoped (electron-doped) high- T_c superconductor $\text{Pr}_{2-x}\text{Ce}_x\text{CuO}_{4-y}$ and the newly synthesized electron-doped material $\text{La}_{2-x}\text{Ce}_x\text{CuO}_{4-y}$ (ref. 8). The existence of a pseudogap when superconductivity is suppressed excludes precursor superconductivity as its origin. Our observation is consistent with the presence of a (quantum) phase transition at $T = 0$, which may be a key to understanding high- T_c superconductivity. This supports the picture that the physics of high- T_c superconductors is determined by the interplay between competing and coexisting ground states.

It is well established that above the superconducting transition

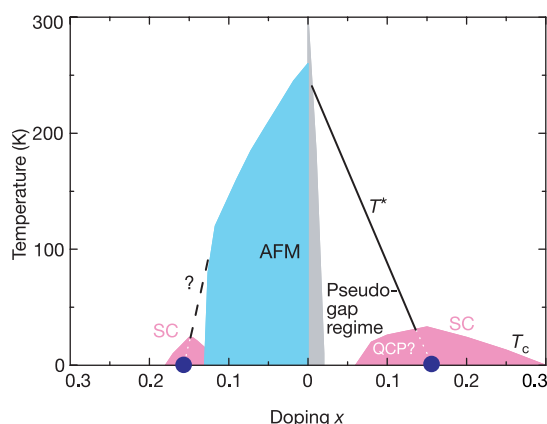


Figure 1 Simplified schematic phase diagram for hole-doped and electron-doped high- T_c superconductors (SC): left, electron-doped ($\text{La}, \text{Pr}, \text{Nd}$) $_{2-x}\text{Ce}_x\text{CuO}_{4-y}$; right, hole-doped $\text{La}_{2-x}\text{Sr}_x\text{CuO}_{4-y}$. The exact behaviour of the characteristic pseudogap temperature $T^*(x)$ is under discussion for the hole-doped case and is almost unknown for the electron-doped side. The different decay of antiferromagnetism (AFM) with increased x is due to spin dilution and spin frustration for electron doping and hole doping, respectively. A possible quantum critical point (QCP) is indicated.

temperature T_c of hole-doped high- T_c superconductors a pseudogap, comparable in size to the superconducting gap, exists in the electronic excitation spectra. The doping-dependent characteristic temperature at which this pseudogap opens is denoted by T^* , which is in the underdoped region significantly larger than T_c (Fig. 1). For recent overviews see refs 1 and 2. However, tunnelling experiments in electron-doped high- T_c superconductors have failed to find such a pseudogap above T_c (ref. 9). Owing to the relatively small values of the upper critical field density in electron-doped high- T_c superconductors, the normal state of these copper oxides can be probed at low temperatures by suppressing superconductivity by an applied field. We have performed tunnelling spectroscopy by using grain boundary junctions as superconductor–insulator–superconductor contacts in epitaxial thin films of the electron-doped superconductors $\text{Pr}_{2-x}\text{Ce}_x\text{CuO}_{4-y}$ and $\text{La}_{2-x}\text{Ce}_x\text{CuO}_{4-y}$. The tunnel junction in this method is composed of a grain boundary formed during epitaxial growth of the material on a bicrystalline substrate. The tunnelling direction is within the a – b plane of the copper oxide superconductors. Because the grain boundary junction is a zig-zag interface¹⁰ this method averages over different grain boundary angles and integrates an area of typically 10^{-10} m^2 . This method therefore does not reveal information on a possible phase separation on a smaller length scale. Experimental details are given in Table 1.

Recently, a suppression of the density of states was found in electron-doped superconductors at an energy scale about 10-fold larger than the superconducting gap by using optical spectroscopy^{11–13}. Here we use the term pseudogap to refer to a different normal-state low-energy-scale excitation gap that is comparable in size to the superconducting gap. This property and the smooth evolution from the superconducting gap below T_c into the pseudogap above T_c in the quasiparticle tunnelling spectrum as observed in hole-doped high- T_c superconductors were interpreted in terms of a precursor superconductivity model¹⁴. Similar experiments were performed for the electron-doped copper oxides $\text{Pr}_{2-x}\text{Ce}_x\text{CuO}_{4-y}$ and $\text{Nd}_{2-x}\text{Ce}_x\text{CuO}_{4-y}$ with the result that, at least for optimum doping and temperatures above T_c , no pseudogap could be detected⁹. However, when the copper oxide superconductor is driven in the normal state by applying a high magnetic field, a clear pseudogap feature at a similar energy scale to the superconducting gap is observed in the quasiparticle tunnelling spectra^{9,15}. The obvious question is how this pseudogap evolves with doping. Figure 2 shows the doping dependence of the pseudogap for two different electron-doped copper oxides measured at 4.2 K in a magnetic field of 10 and 20 T, that is, well above the upper critical field H_{c2} . As can be seen, the pseudogap feature persists up to the highest applied fields and does not depend on the magnetic field. It remains almost unchanged up to 20 T.

Having established the doping-dependent existence of the pseudogap, the next step is to determine the doping dependence of the

characteristic temperature T^* , as shown in Fig. 3 for all dopings. For the investigated doping range, we always find that $T^* < T_c$, in contrast to the hole-doped high- T_c copper oxides, which explains why no pseudogap has been found until now in the electron-doped compounds above T_c . However, the dependence on doping x is qualitatively the same for hole and electron doping: T^* increases approximately linearly with decreased doping.

We address below the possibility that the pseudogap is due to residual superconductivity in a vortex liquid state below a much

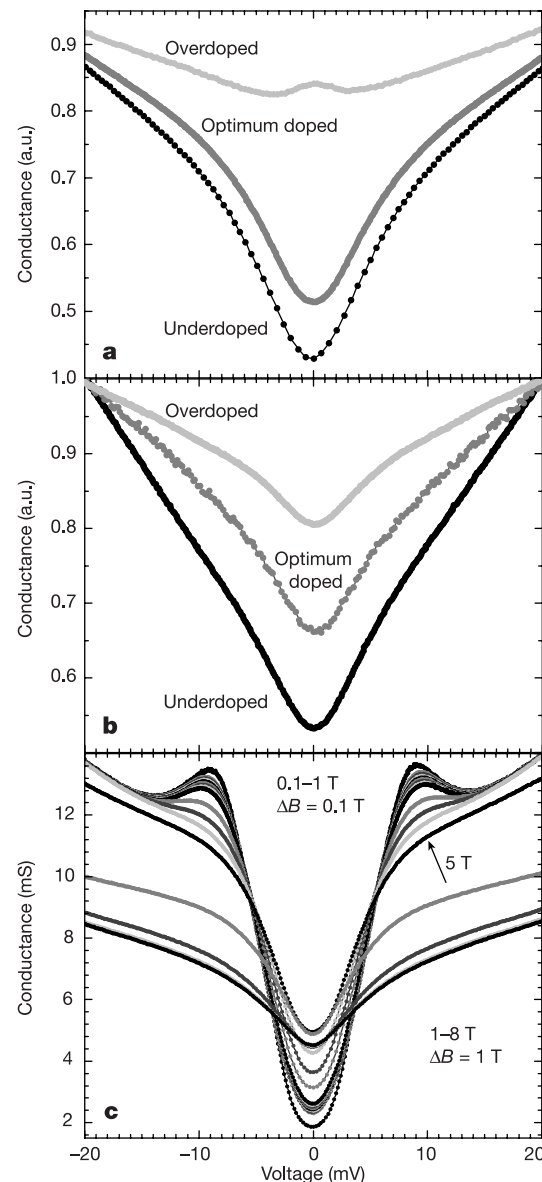


Figure 2 Pseudogap in the magnetic field driven normal state and its doping dependence. **a**, Doping dependence of tunnelling spectra in $\text{La}_{2-x}\text{Ce}_x\text{CuO}_{4-y}$ taken at 10 T and 4.2 K. In the underdoped region there is a clear and well-developed pseudogap structure that becomes smaller with increased doping x . The pseudogap feature vanishes in the overdoped sample. **b**, The same for $\text{Pr}_{2-x}\text{Ce}_x\text{CuO}_{4-y}$ at 20 T and 4.2 K. A small pseudogap survives up to a doping level of $x \approx 0.16$. **c**, Magnetic field series for tunnelling spectra in $\text{La}_{2-x}\text{Ce}_x\text{CuO}_{4-y}$ ($x = 0.116$) at 4.2 K. The coherence peaks at the gap edge of the superconducting gap disappear at H_{c2} , which is between 5 and 6 T. At this field the background conductance drops notably because the electrodes are driven normal. Above 5 T the sample is driven in the normal state, but the pseudogap is present up to the highest applied fields.

Table 1 Doping x and transition temperatures T_c of the thin films used here

$\text{Pr}_{2-x}\text{Ce}_x\text{CuO}_{4-y}$		$\text{La}_{2-x}\text{Ce}_x\text{CuO}_{4-y}$	
x	T_c (K)	x	T_c (K)
0.134 (underdoped)	20	0.098	28
0.148 (optimum doped)	23	0.116	27.5
0.159 (overdoped)	19	0.147	19

The phase diagram of $\text{La}_{2-x}\text{Ce}_x\text{CuO}_{4-y}$ is shifted to lower x values. We used [001] tilt bicrystal SrTiO_3 substrates with a misorientation angle of 36.8° . We did not use $\text{Nd}_{2-x}\text{Ce}_x\text{CuO}_{4-y}$ as the electron-doped copper oxide superconductor, to exclude possible effects from the strong paramagnetism of the Nd^{3+} ion in this compound²⁹. Grain boundary tunnelling spectroscopy has been used successfully to study a – b plane tunnelling in hole-doped high- T_c superconductors revealing, for example, zero-energy bound states due to the intrinsic sign change of the d -wave superconducting pair potential^{29,30}. The values for T_c have been consistently obtained by both resistive measurement of the thin-film electrodes and grain boundary tunnelling. This clearly shows that grain boundary tunnelling reflects the doping-dependent properties of the electrode materials.

higher 'real' H_{c2} . For hole-doped $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ a Nernst signal was detected in the pseudogap phase⁴, indicating the presence of vortices below T^* but above T_c . For electron-doped superconductors the transport entropy of the magnetic flux line also indicates a larger H_{c2} than determined from resistivity measurements¹⁶. However, for electron doping the increase of H_{c2} is only of the order of a few teslas at maximum and therefore cannot account for the observed gap here, which persists even well above 20 T.

For the discussion of our results, we summarize the doping dependence of T^* and T_c for $\text{Pr}_{2-x}\text{Ce}_x\text{CuO}_{4-y}$ in a schematic phase diagram in Fig. 4. T_c retraces the dome-shaped superconducting phase shown in Fig. 1. The pseudogap regime characterized by T^* has a similar doping dependence to that of the hole-doped copper oxides with the important difference that $T^* < T_c$

in the investigated doping range. From extrapolation of the data, we predict that below $x \approx 0.13$ a pseudogap will also be observed above T_c , as occurs with hole-doped high- T_c superconductors at higher doping. In contrast, going into the overdoped region the $T^*(x)$ curve seems to vanish at $x \approx 0.16$ – 0.18 , which could mark a critical doping level x_c . We note that for hole-doped $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+x}$ a similar observation to that in our experiment has been reported from c -axis intrinsic tunnelling spectroscopy: the coexistence of a superconducting gap dependent on temperature and magnetic field, and a field-independent pseudogap feature¹⁷. The observed behaviour might therefore be universal in the copper oxide phase diagram. In a different experiment¹⁸, the temperature dependence of the resistivity, $\rho(T)$, as a function of doping for the electron-doped $\text{Pr}_{2-x}\text{Ce}_x\text{CuO}_4$ revealed a metal–insulator crossover around a critical doping of $x \approx 0.15$, which is consistent with our results. We note that $\rho(T)$ behaves similarly in hole-doped $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$, again underlining the generality of the phenomenon in high- T_c copper oxides¹⁹.

We now address the important question of whether the pseudogap regime coexists with superconductivity or is established only when superconductivity is destroyed by the magnetic field. From our data it seems that the pseudogap is 'hidden' below the coexisting global superconducting phase. First, the tunnelling conductance evolves smoothly at the Fermi level as a function of the applied field with no sign of an opening of a pseudogap in the superconducting spectra, as shown in Fig. 2c. This indicates that the pseudogap already underlies the spectra at zero field. Second, the coherence peaks in the superconducting state are more suppressed in the underdoped region than in the optimum and overdoped case owing to the presence of the pseudogap, suggesting a competition between two order parameters. Third, the conservation rule for density of states is recovered only after removing the pseudogap spectral weight. This again shows that the pseudogap is already present at zero applied field.

It is a general feature of the copper oxide high- T_c superconductors that because of strong electronic correlations there are many more than only two phases with a transition from a normal Fermi liquid to the superconducting state. The key result of our experiments is to establish the existence of a T^* line within the superconducting region in the copper oxide phase diagram. A relation between the observed pseudogap and precursor superconductivity can therefore be excluded immediately. It is tempting to identify the T^* line with an additional phase boundary due to a second order

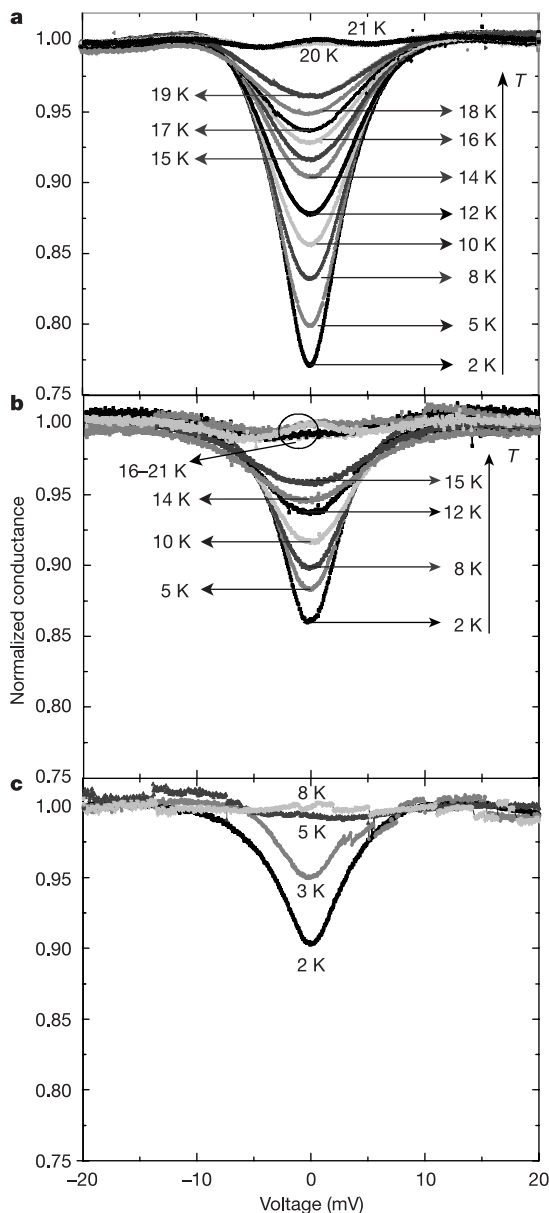


Figure 3 Temperature dependence of the pseudogap in high magnetic fields. Normalized spectra at different temperatures are shown for all investigated dopings of $\text{Pr}_{2-x}\text{Ce}_x\text{CuO}_{4-y}$ in a field of 14 T: **a**, underdoped; **b**, optimum doping; **c**, overdoped. The suppression of the density of states at the Fermi energy becomes smoothly weaker with increasing temperature and vanishes suddenly at a temperature we define as $T^*(x)$.

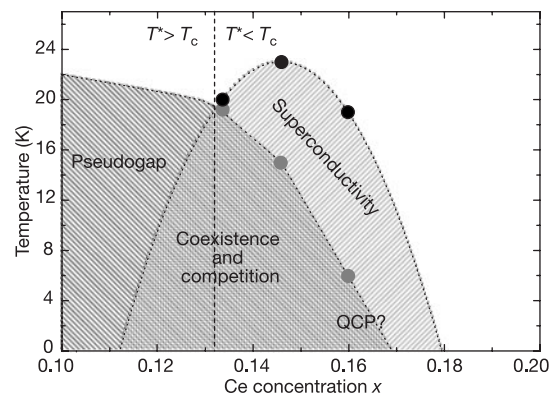


Figure 4 Phase diagram of the electron-doped high- T_c superconductor $\text{Pr}_{2-x}\text{Ce}_x\text{CuO}_{4-y}$. The data reveal the coexistence of superconductivity and a pseudogap regime in a wide doping range: black data points, T_c ; grey data points, T^* ; the point size corresponds to the error. The curves through the points show a possible continuation of the phase diagram. Extrapolating $T^*(x)$ suggests a quantum critical point in the overdoped regime at $0.16 < x < 0.18$. A possible quantum critical point (QCP) is indicated.

parameter hidden in the copper oxide phase diagram. Recent angle-resolved photoemission spectroscopy (ARPES) measurements indicate a possible broken time-reversal symmetry in the pseudogap regime, implying that it is a real thermodynamic phase²⁰; this must be established by further experiments. This ARPES experiment has been interpreted as support for a phase with circulating currents⁵. Indeed, several theoretical studies on hole-doped high- T_c superconductors have highlighted the possibility of a quantum critical point in the phase diagram involving the existence of a second order parameter^{1,5–7,21–23}. It is already well known that antiferromagnetic Néel order and a charge/spin (stripe) density wave exist in the copper oxide phase diagram. In addition, the possibility of a d -density wave⁷ (staggered flux) order and $d_{x^2-y^2} + id_{xy}$ or $d_{x^2-y^2} + is$ superconductivity has been discussed in detail. Recent neutron-scattering experiments on $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ (ref. 24) and $\text{YBa}_2\text{Cu}_3\text{O}_{6.6}$ (ref. 25) have revealed enhanced antiferromagnetic correlations in the underdoped superconducting region. Another example of hidden order was found in the heavy fermion metal URu_2Si_2 (ref. 26). In the phase diagram suggested here for electron-doped high- T_c superconductors, there (co)exist at least two competing order parameters. A small applied magnetic field couples only to the superconducting order parameter, suppressing it completely at H_{c2} leaving the pseudogap-related order parameter at a finite value, as predicted recently for a d -density-wave ordered state²⁷. This excludes the possibility that the pseudogap regime is related to a second superconducting order parameter with different symmetry²⁸. Owing to the competition of the two order parameters, the magnetic field needed to destroy superconductivity can indeed be smaller than H_{c2} because of the additional suppression of superconductivity by the competing order. Strong interactions between the two order parameters can result in even more complicated phase diagrams with more phases than shown in Fig. 4. The identification of the possible order in the pseudogap phase is one of the most prominent questions in the field of high- T_c superconductivity. Any suggested order should be consistent with the observations that the pseudogap coexists with superconductivity and is essentially unchanged by a large applied external field. In particular, these constraints exclude the physical picture of precursor superconductivity. □

Received 10 July 2002; accepted 7 February 2003; doi:10.1038/nature01488.

1. Tallon, J. L. & Loram, J. W. The doping dependence of T^* —what is the real high- T_c phase diagram? *Physica C* **349**, 53–68 (2001).
2. Timusk, T. & Statt, B. The pseudogap in high-temperature superconductors: an experimental survey. *Rep. Prog. Phys.* **62**, 61–122 (1999).
3. Emery, V. J. & Kivelson, S. A. Importance of phase fluctuations in superconductors with small superfluid density. *Nature* **374**, 434–437 (1995).
4. Xu, Z. A., Ong, N. P., Wang, Y., Kakeshita, T. & Uchida, S. Vortex-like excitations and the onset of superconducting phase fluctuation in underdoped $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$. *Nature* **406**, 486–488 (2000).
5. Varma, C. M. Pseudogap phase and the quantum-critical point in copper-oxide metals. *Phys. Rev. Lett.* **83**, 3538–3541 (1999).
6. Sachdev, S. Quantum criticality: competing ground states in low dimensions. *Science* **288**, 475–480 (2000).
7. Chakravarty, S., Laughlin, R. B., Morr, D. K. & Nayak, C. Hidden order in the cuprates. *Phys. Rev. B* **63**, 094503 (2001).
8. Naito, M. & Hepp, M. Superconducting T^* - $\text{La}_{2-x}\text{Ce}_x\text{CuO}_4$ films grown by molecular beam epitaxy. *Jpn. J. Appl. Phys.* **39**, L485–L487 (2000).
9. Kleefisch, S. *et al.* Possible pseudogap behavior of electron-doped high-temperature superconductors. *Phys. Rev. B* **63**, R100507 (2001).
10. Mannhart, J. *et al.* Generation of magnetic flux by single grain boundaries of $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$. *Phys. Rev. Lett.* **77**, 2782–2785 (1996).
11. Onose, Y., Taguchi, Y., Ishizaka, K. & Tokura, Y. Doping dependence of pseudogap and related charge dynamics in $\text{Nd}_{2-x}\text{Ce}_x\text{CuO}_4$. *Phys. Rev. Lett.* **87**, 217001 (2001).
12. Singley, E. J., Basov, D. N., Kurauchi, K., Uefuji, T. & Yamada, K. Electron dynamics in $\text{Nd}_{1.85}\text{Ce}_{0.15}\text{CuO}_{4+\delta}$: Evidence for the pseudogap state and unconventional c -axis response. *Phys. Rev. B* **64**, 224503 (2001).
13. Armitage, N. P. *et al.* Anomalous electronic structure and pseudogap effects in $\text{Nd}_{1.85}\text{Ce}_{0.15}\text{CuO}_4$. *Phys. Rev. Lett.* **87**, 147003 (2001).
14. Renner, Ch., Revaz, B., Genoud, J.-Y., Kadowaki, K. & Fischer, Ø. Pseudogap precursor of the superconducting gap in under- and overdoped $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. *Phys. Rev. Lett.* **80**, 149–152 (1998).
15. Biswas, A. *et al.* Gapped tunneling spectra in the normal state of $\text{Pr}_{2-x}\text{Ce}_x\text{CuO}_4$. *Phys. Rev. B* **64**, 104519 (2001).
16. Gollnik, F. & Naito, M. Doping dependence of normal- and superconducting-state transport properties of $\text{Nd}_{2-x}\text{Ce}_x\text{CuO}_{4+y}$ thin films. *Phys. Rev. B* **58**, 11734–11752 (1998).

17. Krasnov, V. M., Yurgens, A., Winkler, D., Delsing, P. & Claeson, T. Evidence for coexistence of the superconducting gap and the pseudogap in Bi-2212 from intrinsic tunneling spectroscopy. *Phys. Rev. Lett.* **84**, 5860–5863 (2000).
18. Fournier, P. *et al.* Insulator-metal crossover near optimal doping in $\text{Pr}_{2-x}\text{Ce}_x\text{CuO}_4$: anomalous normal-state low temperature resistivity. *Phys. Rev. Lett.* **81**, 4720–4723 (1998).
19. Ando, Y., Boebinger, G. S., Passner, A., Kimura, T. & Kishio, K. Logarithmic divergence of both in-plane and out-of-plane normal-state resistivities of superconducting $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ in the zero-temperature limit. *Phys. Rev. Lett.* **75**, 4662–4665 (1995).
20. Kaminski, A. *et al.* Spontaneous breaking of time-reversal symmetry in the pseudogap state of a high- T_c superconductor. *Nature* **416**, 610–613 (2002).
21. Varma, C. M. Proposal for an experiment to test a theory of high-temperature superconductors. *Phys. Rev. B* **61**, R3804–R3807 (2000).
22. Andergassen, S., Caprara, S., Di Castro, C. & Grilli, M. Anomalous isotopic effect near the charge-ordering quantum criticality. *Phys. Rev. Lett.* **87**, 056401 (2001).
23. Manske, D., Dahm, T. & Bennemann, K.-H. Phase diagram of hole-doped high- T_c superconductors: Effects of Cooper-pair phase fluctuations within fluctuation-exchange theory. *Phys. Rev. B* **64**, 144520 (2001).
24. Lake, B. *et al.* Antiferromagnetic order induced by an applied magnetic field in a high-temperature superconductor. *Nature* **415**, 299–302 (2002).
25. Mook, H. A., Dai, P. & Dogan, F. Observation of magnetic moments in the superconducting state of $\text{YBa}_2\text{Cu}_3\text{O}_{6.6}$. *Phys. Rev. B* **64**, 012502 (2001).
26. Chandra, P., Coleman, P., Mydosh, J. A. & Tripathi, V. Hidden orbital order in the heavy fermion metal URu_2Si_2 . *Nature* **417**, 831–834 (2002).
27. Nguyen, H. K. & Chakravarty, S. Effects of magnetic field on the d -density-wave order in the cuprates. *Phys. Rev. B* **65**, R180519 (2002).
28. Skinta, J. A., Kim, M.-S., Lemberger, T. R., Greibe, T. & Naito, M. Evidence for a transition in the pairing symmetry of the electron-doped cuprates $\text{La}_{2-x}\text{Ce}_x\text{CuO}_{4-y}$ and $\text{Pr}_{2-x}\text{Ce}_x\text{CuO}_{4-y}$. *Phys. Rev. Lett.* **88**, 207005 (2002).
29. Alff, L. *et al.* Anomalous low temperature behavior of superconducting $\text{Nd}_{1.85}\text{Ce}_{0.15}\text{CuO}_{4-y}$. *Phys. Rev. Lett.* **83**, 2644–2647 (1999).
30. Alff, L. *et al.* Observation of bound surface states in grain-boundary junctions of high-temperature superconductors. *Phys. Rev. B* **58**, 11197–11200 (1998).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to L.A. (e-mail: lambert.alff@wmi.badw.de).

An inverse transition of magnetic domain patterns in ultrathin films

O. Portmann, A. Vaterlaus & D. Pescia

Laboratorium für Festkörperphysik, Eidgenössische Technische Hochschule Zürich, CH-8093 Zürich, Switzerland

Inverse freezing and inverse melting are processes where a more symmetric phase is found at lower temperatures than at higher temperatures. Such inverse transitions are very rare¹. Here we report the existence of an inverse transition effect in ultrathin Fe films that are magnetized perpendicular to the film plane. The magnetization of these films is not uniform, but instead manifests itself as stripe domains with opposite perpendicular magnetization^{2–4}. Predictions relating to the disordering of this striped ground state in the limit of monolayer film thicknesses are controversial. Mean-field arguments^{5–7} predict a continuous reduction of the stripe width when the temperature is increased; other studies^{8–11} suggest that topological defects, such as dislocations and disclinations, might penetrate the system and induce geometrical phase transitions. We find, from scanning electron microscopy imaging, that when the temperature is increased, the low-temperature stripe domain structure transforms into a more symmetric, labyrinthine structure. However, at even higher temperatures and before the loss of magnetic order, a re-occurrence of the less symmetric stripe phase is found. Despite the widespread theoretical and experimental work on striped systems, this phase sequence and the microscopic instabilities driving it have not been observed before.

To image the domain structures in face-centred-cubic (f.c.c.) Fe films on Cu(001), we use scanning electron microscopy with polarization analysis (SEMPA; see ref. 12 for details). The black-and-white contrast in the images corresponds to domains of opposite perpendicular magnetization. In order to emphasize details of the images, we sometimes colour white stripes yellow, and dark stripes red or blue. No magnetic field has been applied in this work¹².

The evolution of the domain pattern at room temperature as the film thickness is decreased is shown in Fig. 1a–d. Decreasing the film thickness is equivalent to increasing the effective temperature of the film (Fig. 2). Four distinct patterns can be recognized: a stripe pattern at low effective temperatures, followed by a labyrinthine pattern, a re-entrant stripe domain structure, and finally a contrastless, paramagnetic phase. In Fig. 1e–h the same sequence is observed at a fixed location on a film when the temperature is increased. The inverse freezing process—that is, the transformation from the more symmetric labyrinthine phase to the less symmetric stripe phase—is seen in Fig. 1b, c and Fig. 1f, g. In Fig. 1i–k the temperature is lowered, and again a fixed film location is imaged. The stripe phase (Fig. 1j) emerges from the paramagnetic phase (Fig. 1i) and undergoes inverse melting to a labyrinthine phase (Fig. 1k). This process is extremely slow (Fig. 1k was taken about 12 hours after Fig. 1j). The slow kinetics are probably due to the necessity of rearranging long stripes in a situation of decreasing temperature. Thus, the next transition—from the labyrinthine to the

low-temperature stripe phase—is very rare, although occasionally observed (Fig. 1l, m).

In Fig. 2 we report some quantitative aspects of the domain structure. Films grown on substrates held at about room temperature during evaporation have a thickness-dependent Curie temperature T_C (refs 13–15) (here substrate temperature is about 315 K, and deposition rate is about 0.1 monolayers per minute, ML min^{-1}). At low coverages, T_C increases with thickness (Fig. 2a). In this thickness range, decreasing the thickness at constant temperature drives the film towards the phase transition, and is thus equivalent to establishing an increasing effective temperature. After reaching a maximum value, T_C decreases again (Fig. 2a). The value of the maximum and the thickness range for which T_C is above room temperature depend on the growing conditions and on how T_C is determined^{13,15,16}. What produces the decrease of T_C is still a controversial point, and is beyond the scope of this Letter.

Mean-field theory^{4–7}, which does not consider geometrical phase transitions, predicts the width L of the domains to decrease continuously when the temperature or the effective temperature are raised. The stripes assume a finite width L_C at T_C where the magnetization within each stripe vanishes. L_C at T_C was calculated in ref. 6 to be about $4\pi\Gamma/\Omega$, where Γ is a measure of the exchange interaction and Ωa is the strength of the dipolar interaction (here a is the in-plane nearest-neighbour distance). Theoretical work^{2,4–6} shows that in a situation of perpendicular magnetization, the value of the magnetic anisotropy does not appear explicitly in the expression for L_C and that L is determined, close to T_C , by the ratio of the exchange to the dipolar interaction. The perpendicular magnetic anisotropy, determined for instance in ref. 17, is a smoothly varying function of the temperature. Setting $a = 0.25 \text{ nm}$, Γ to $T_C \approx 300 \text{ K}$ and the strength of the dipolar interaction Ωa to $\sim 2.5 \text{ K}$, we calculate $L_C \approx 300 \text{ nm}$. Figure 2b confirms this value. While imaging from the top to the bottom, the film cools down from the paramagnetic state (top of Fig. 2b), followed by a sudden appearance of stripes with a finite width of $\sim 500 \text{ nm}$. Figure 2c shows that the mean stripe width increases

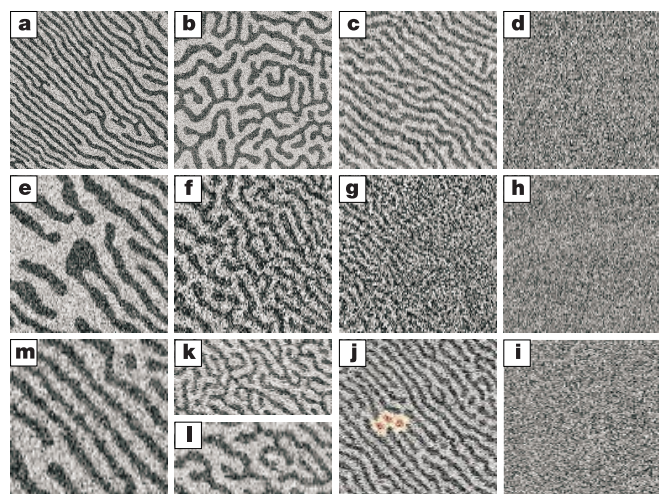


Figure 1 The four phases in ultrathin f.c.c. Fe films. **a–d**, The effective temperature is increased from left to right. The paramagnetic phase (**d**) appears contrastless. (By contrastless we mean that the spin polarization averages to zero within 10 ms, which is the measuring time per pixel.) For each image, the image width β and the film parameters δ (thickness) and T (temperature) are given. If not specified, the vertical scale of the images is the same as the horizontal scale. Slightly varying quantities are indicated by the $\approx$ sign. ($\beta_a = 92 \mu\text{m}$, $\beta_b = 46 \mu\text{m}$, $\beta_c = 23 \mu\text{m}$, $\beta_d = 23 \mu\text{m}$, $\delta_a = 2.1 \text{ ML}$, $\delta_b = 1.91 \text{ ML}$, $\delta_c = 1.84 \text{ ML}$, $\delta_d = 1.73 \text{ ML}$, $T = 293 \text{ K}$.) **e–h**, A fixed region of a film is imaged while the temperature is increased from **e** to **h**. The white stripes in the low-temperature phase are slightly broader than the black stripes: this is due to some residual, unavoidable small magnetic field, as discussed in ref. 6. ($\beta_e = 46 \mu\text{m}$, $\beta_f = 23 \mu\text{m}$, $\beta_g = 23 \mu\text{m}$, $\beta_h = 23 \mu\text{m}$, $\delta \approx 1.89 \text{ ML}$, $T_e \approx 210 \text{ K}$, $T_f \approx 268 \text{ K}$, $T_g \approx 283 \text{ K}$, $T_h \approx 313 \text{ K}$.) **i–k**, A fixed region of a film is imaged while the temperature is lowered from **i** (paramagnetic phase) to **k** (labyrinthine phase). Notice the bubble-like defects (**j**, red), which are known in pattern formation to be possible precursors of a labyrinthine pattern²³. ($\beta_i = 23 \mu\text{m}$, $\beta_j = 44 \mu\text{m}$, $\beta_k = 44 \mu\text{m}$, $\delta = 1.97 \text{ ML}$, $T_i \approx 321 \text{ K}$, $T_j \approx 300 \text{ K}$, $T_k \approx 296 \text{ K}$.) **l–m**, The labyrinthine phase transforms into the stripe pattern on cooling. ($\beta_l = 33 \mu\text{m}$, $\beta_m = 33 \mu\text{m}$, $\delta \approx 1.94 \text{ ML}$, $T_l \approx 294 \text{ K}$, $T_m \approx 281 \text{ K}$.)

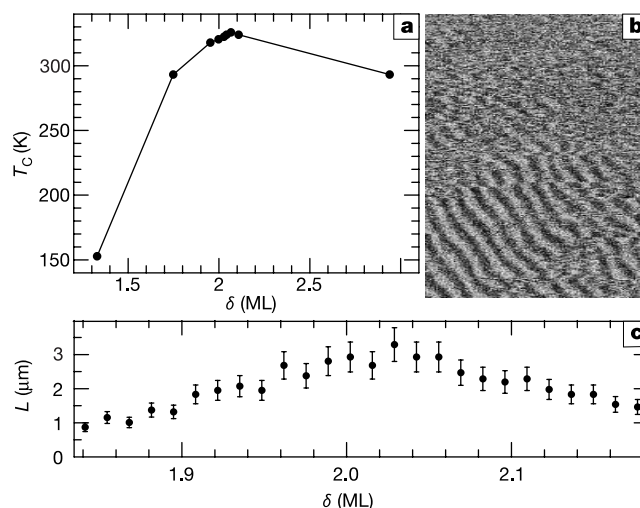


Figure 2 Mean-field aspects of the phase transition. **a**, The phase transition line. Above the line the system is paramagnetic, below it the films are magnetized perpendicular to the film plane. **b**, The finite stripe width at T_C . The film cools down during image taking from the paramagnetic phase (top of the image) to the stripe phase (bottom). ($\beta = 14 \mu\text{m}$, $\delta = 1.94 \text{ ML}$.) **c**, Mean stripe width at room temperature as a function of the film thickness. The effective temperature is lowest in the centre and increases towards left and right.

smoothly when the effective temperature is lowered, as expected from mean-field arguments. Notice that the maximum of the transition line of Fig. 2a is mirrored in the curve in Fig. 2c, which also undergoes a maximum.

We now look at the role of dislocations. Stripe patterns contain four types of dislocations¹² (shown coloured in Fig. 3a). Most dislocations appear in pairs, possibly separated by one or two continuous stripes: the stripes form the 'smectic-like' phase predicted to develop from the ground state^{9–11,18}. The structure factor (inset), obtained by Fourier transformation of the image, has two well-defined spots along one direction in reciprocal space: that is, this stripe phase is orientationally ordered¹⁰. Imaging the stripes on a wedge (Fig. 3b, the effective temperature increases from left to right) demonstrates the role of dislocations. They allow for new stripes to be inserted (shown coloured) so that the stripe density can increase when T_C is approached, in apparent agreement with mean-field arguments⁵. Notice, however, that this model predicts a continuous adjustment of the stripe width, which is only observed if the mean stripe width is considered.

When the temperature is increased at uniform thickness the dislocations are also seen to glide, the net result being again an increase of the stripe density. This process appears to be forbidden in thick films, which prefer to form a labyrinthine pattern instead¹⁹. The analogous process of stripe annihilation is shown in Fig. 3c. The stripe pattern is imaged from top to bottom while emerging from the uniform paramagnetic state upon cooling. The three yellow stripes on the top collapse into two by a sudden annihilation process accompanied by a rapid zig-zag movement of the neighbouring

stripes (also shown coloured in the figure)—first they move to the left, then back to the right. This movement allows for the adjustment of the change in periodicity produced by stripe annihilation. As thinner stripes merge through a number of discrete steps, the number of stripes within a given field of view is reduced (as can be

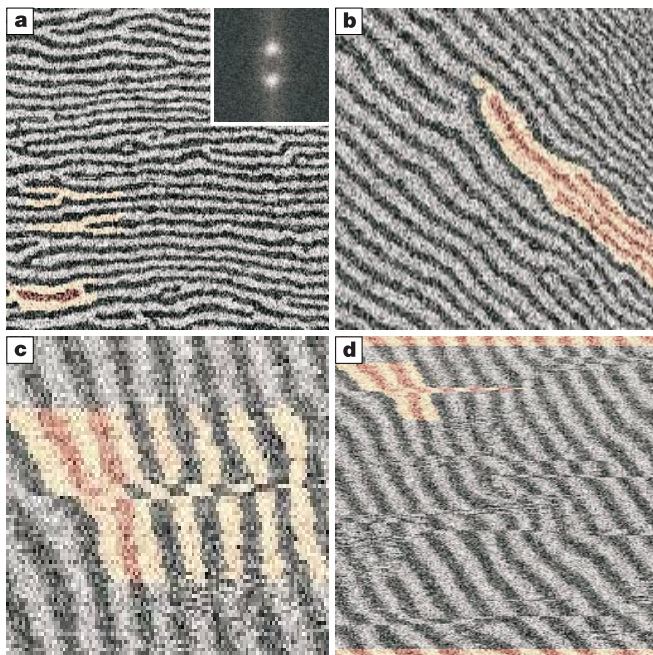


Figure 3 The role of dislocations. **a**, The low-temperature stripe phase, with four types of defects (shown coloured). Inset: structure factor, indicating that the stripes have a local directional order. ($\beta = 46 \mu\text{m}$, $\delta = 2.44 \text{ ML}$, $T = 293 \text{ K}$.) **b**, The effective temperature increases from left to right and the stripe density increases by insertion of new stripes (shown coloured). ($\beta = 65 \mu\text{m}$, δ from 2.33 ML to 2.12 ML, $T = 286 \text{ K}$.) **c**, **d**, The images were taken when the sample was cooling down from the paramagnetic phase to the high-temperature stripe phase. The top of the images has the highest temperature. The stripe density decreases on cooling by annihilation of stripes, followed by the rearrangement of the remaining stripes (shown coloured). ($\beta_c = 12 \mu\text{m}$ (vertical size, $6 \mu\text{m}$), $\beta_d = 18.4 \mu\text{m}$, $\delta = 1.96 \text{ ML}$, $T \approx 295 \text{ K}$.)

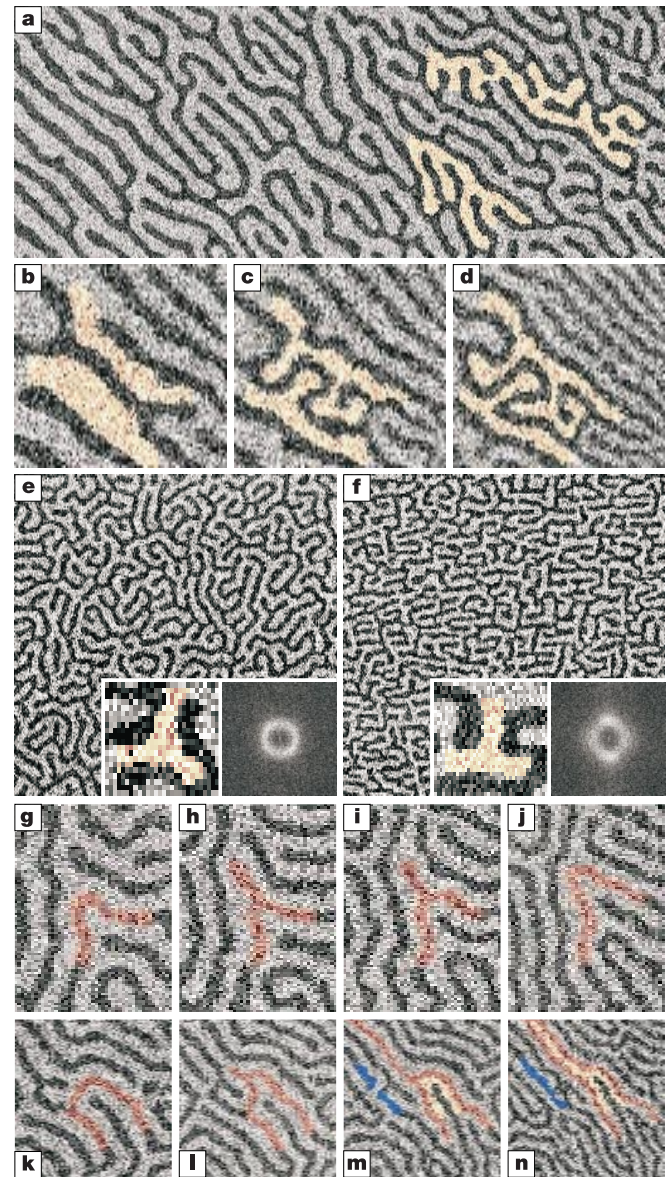


Figure 4 Microscopic mechanisms of transformations. In **a–f** the formation of the labyrinthine pattern is illustrated, in **g–n** the inverse transformation from a labyrinthine to a high-temperature stripe pattern is imaged. **a**, The effective temperature increases from left to right. The emission of transversal 'fingers' is coloured. ($\beta = 155 \mu\text{m}$, δ varies from 2.06 ML to 1.98 ML, $T = 294 \text{ K}$.) **b–d**, The temperature is increased from left to right for the same film region. The stripes are seen to emit fingers (shown coloured) that wind up to promote the formation of the labyrinthine pattern. ($\beta = 36 \mu\text{m}$, $\delta \approx 2.19 \text{ ML}$, $T_b \approx 298 \text{ K}$, $T_c \approx 304 \text{ K}$ and $T_d \approx 305 \text{ K}$.) **e**, The labyrinthine pattern. The insets show a triangular V-disclination and the structure factor. ($\beta = 92 \mu\text{m}$, $\delta = 1.97 \text{ ML}$, $T = 293 \text{ K}$.) **f**, The labyrinthine pattern on a stepped substrate. The disclination has a square symmetry, as does the structure factor. (Parameters as in **e**, $\delta = 2.15 \text{ ML}$.) **g–j** and **k–n**, Knee-bend and bridge instabilities (shown coloured) leading to the straightening of the labyrinthine pattern when the temperature is increased (from left to right). ($\beta_{g-j} = 23 \mu\text{m}$, $\beta_{k-n} = 33 \mu\text{m}$, $\delta = 1.95 \text{ ML}$, T is raised from $T_g \approx 311 \text{ K}$ to $T_j \approx 315 \text{ K}$ and from $T_k \approx 311 \text{ K}$ to $T_n \approx 316 \text{ K}$.)

seen by counting the number of stripes at the top (the 'hot side' of the image) and at the bottom (the 'cold side') of Fig. 3d). Notice that a sudden movement of the sample can be excluded, as stripes not involved in the adjustment are not affected by any discontinuity.

We now consider the microscopic mechanisms of transformations. Figure 4a shows a stripe pattern on a wedge. The effective temperature increases from left to right. The stripes emit fingers and branches transversally to themselves that wind up to initiate orientational melting (some examples are coloured in the figure). The same process is observed when the temperature is increased (Fig. 4b–d). This transversal front instability, predicted in a different context²⁰, promotes the loss of orientational order in favour of a 'labyrinthine' pattern (Fig. 4e)—a connected network of domains containing two types of defects: convex disclinations (X-type, a domain ending as a tip surrounded by a domain with opposite magnetization looping around it) and concave disclinations (V-type, three segments meeting at one point, see Fig. 4e inset)²¹. A surprising feature is the angle formed by the segments in V-disclinations—predominantly 120°. This local three-fold symmetry can be understood within a planar isotropic system as being due to the repulsive dipolar interaction between segments of the same 'colour', but is in striking contrast to the square symmetry imposed on the lattice by the epitaxial growth on the (100) face, as confirmed by our low-energy electron diffraction images.

The structure factor of the labyrinthine phase (Fig. 4e, inset) has its spectral weight on a ring that is weakly hexagonally shaped, confirming that the orientationally melted state consists of disclinations with three-fold symmetry that have retained a memory of the original stripe orientation. This 'rotational incommensurability' indicates that the domain structure, residing on a micrometre scale, is totally disjoint from the lattice geometry, residing on $a \approx 0.1$ nm scale. We conclude that the films indeed follow the 'floating solid' hypothesis^{5,9–11} formulated for two-dimensional striped systems. The square symmetry is restored when the Fe film is grown on top of a slightly miscut Cu(100) surface. The miscut ($\sim 3^\circ$) creates an array of oriented monatomic steps within the substrate (periodicity ~ 3.5 nm). The stripes in films grown on top of the miscut substrate preferentially align parallel or perpendicular to the steps. The system of oriented monatomic steps does not prevent orientational melting, but the labyrinthine phase emerging from the stripes (Fig. 4f) contains the 90° corners (inset) predicted by theoretical work on a square lattice^{9,10}. Accordingly, the spectral weight of the structure factor resides on a square-shaped ring, with maxima at its corners (inset) and resembles the structure factor obtained from Monte Carlo simulations in ref. 10.

These observations clarify the role of the substrate: it provides pinning centres so that the stripe positions are not allowed to fluctuate in time as required for a truly 'floating solid'⁵, but does not change the nature of the orientational melting process. This pinning is also essential for selecting locally a direction along which stripes are oriented, although the local direction might vary over large distances, as discussed in ref. 12. In line with the standard model of two-dimensional melting^{8,9}, we observe the intervening of an intermediate labyrinthine pattern before paramagnetism sets in. Notice, however, that the formation of this intermediate phase is promoted by a transversal front instability of the stripes, and not by a standard decay of dislocations into unbound disclinations¹¹.

The main microscopic mechanisms of the transition from the labyrinthine pattern to the re-entrant stripe pattern are coloured in Fig. 4g–n. The knee-bend (Fig. 4g, red) emits a segment (Fig. 4h, red) into the adjacent domain of opposite magnetization. The resulting V-disclination opens to a bell-shaped domain (Fig. 4i, red) and transforms into a knee-bend (Fig. 4j, red) again, the main result being the straightening of the original knee-bend area. A similar straightening is produced by a segment emitted by a knee-

bend (Fig. 4k–l, red), proceeding into the domain of opposite magnetization and cutting it as a pair of scissors (Fig. 4l, m, red, and Fig. 4m, n, yellow). This mechanism simultaneously allows for the required increase of the stripe density. The emission of segments at knee-bends is reminiscent of the knee-bend instability against formation of VX pairs of disclinations observed as solutions of the Cross-Newell pattern formation equation²¹. Segments are preferentially emitted into one direction, along which the re-entrant stripes align. In a third mechanism, two segments separated by a bridge (Fig. 4m, blue) join into a stripe (Fig. 4n). This mechanism removes dislocations in the re-entrant stripe phase. Reordering neither proceeds via an 'inverted' transversal instability nor via the rebinding of disclinations into dislocations as reported in a similar—although non-two-dimensional melting—labyrinthine-stripe transformation²².

The present work reports some unpredicted geometrical features of the magnetic phase transition in ultrathin f.c.c. Fe films. These unusual elements should be significant for any type of stripe order, even if it is encountered in a different context, such as the striped quantum liquid crystal phases envisaged recently in connection with copper oxide superconductors¹⁸. We note that the features reported here are also missing from work on striped quantum liquids. □

Received 23 September 2002; accepted 21 February 2003; doi:10.1038/nature01538.

- Greer, A. L. Too hot to melt. *Nature* **404**, 134–135 (2000).
- Garel, T. & Doniach, S. Phase transitions with spontaneous modulation—the dipolar Ising ferromagnet. *Phys. Rev. B* **26**, 325–329 (1982).
- Seul, M. & Andelman, D. Domain shapes and patterns: the phenomenology of modulated phases. *Science* **267**, 476–483 (1995).
- Sornette, D. Undulation instability in stripe domain structures of bubble materials. *J. Phys.* **48**, 151–163 (1987).
- Czech, R. & Villain, J. Instability of two-dimensional Ising ferromagnets with dipole interaction. *J. Phys. Condens. Matter* **1**, 619–627 (1989).
- Kashuba, A. & Pokrovsky, V. L. Stripe domain structures in a thin ferromagnetic film. *Phys. Rev. B* **48**, 10335–10344 (1993).
- Gehring, G. A. & Keskin, M. The temperature dependence of the domain spacing in ultrathin magnetic films. *J. Phys. Condens. Matter* **5**, L581–L585 (1993).
- Nelson, D. R. in *Fundamental Problems in Statistical Mechanics V* (ed. Cohen, E. G. D.) 53–108 (North-Holland, Amsterdam, 1980).
- Abanov, A., Kalatsky, V., Pokrovsky, V. L. & Saslow, W. M. Phase diagram of ultrathin ferromagnetic films with perpendicular anisotropy. *Phys. Rev. B* **51**, 1023–1038 (1995).
- De'Bell, K., MacIsaac, A. B. & Whitehead, J. P. Dipolar effects in magnetic thin films and quasi-two-dimensional systems. *Rev. Mod. Phys.* **72**, 225–257 (2000).
- Stoycheva, A. D. & Singer, S. J. Stripe melting in a two-dimensional system with competing interactions. *Phys. Rev. Lett.* **84**, 4657–4660 (2000).
- Vaterlaus, A. et al. Two-step disordering of perpendicularly magnetized ultrathin films. *Phys. Rev. Lett.* **84**, 2247–2250 (2000).
- Stampanoni, M., Vaterlaus, A., Aeschlimann, M., Meier, F. & Pescia, D. Magnetic properties of thin fcc iron films on Cu(001). *J. Appl. Phys.* **64**, 5321–5324 (1988).
- Thomassen, J. et al. Magnetic live surface layers in Fe/Cu(100). *Phys. Rev. Lett.* **69**, 3831–3834 (1992).
- Kirilyuk, A., Giergiel, J., Shen, J., Straub, M. & Kirschner, J. Growth of stabilized γ -Fe films and their magnetic properties. *Phys. Rev. B* **54**, 1050–1063 (1996).
- Man, K. L., Altman, M. S. & Poppa, H. Spin polarized low energy electron microscopy investigations of magnetic transitions in Fe/Cu(100). *Surf. Sci.* **480**, 163–172 (2001).
- Pierce, J. P., Torija, M. A., Shen, J. & Plummer, E. W. Mapping the magnetic phase diagram of metastable fct Fe/Cu(100) using Co atoms. *Phys. Rev. B* **64**, 224409 (2001).
- Kivelson, S. A., Fradkin, E. & Emery, V. J. Electronic liquid-crystal phases of a doped Mott insulator. *Nature* **393**, 550–553 (1998).
- Seul, M. & Wolfe, R. Evolution of disorder in two-dimensional stripe patterns: Smectic instabilities and dislocation unbinding. *Phys. Rev. Lett.* **68**, 2460–2463 (1992).
- Yochelis, A., Hagberg, A., Meron, E., Lin, A. L. & Swinney, H. L. Development of standing-wave labyrinthine patterns. *SIAM J. Appl. Dyn. Syst.* **1**, 236–247 (2002).
- Newell, A. C., Passot, T., Bowman, C., Ercolani, N. & Indik, R. Defects are weak and self-dual solutions of the Cross-Newell phase diffusion equation for natural patterns. *Physica D* **97**, 185–205 (1996).
- Harrison, C. et al. Mechanism of ordering in striped patterns. *Science* **290**, 1558–1560 (2000).
- Pearson, J. E. Complex patterns in a simple system. *Science* **261**, 189–192 (1993).

Acknowledgements We thank the Swiss National Foundation and ETH Zurich for financial support.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to O.P. (e-mail: portmann@solid.phys.ethz.ch).

Dissociative hydrogen adsorption on palladium requires aggregates of three or more vacancies

T. Mitsui*, M. K. Rose*†, E. Fomin*†, D. F. Ogletree* & M. Salmeron*

* Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

† Department of Physics, University of California, Berkeley, California 94720, USA

During reaction, a catalyst surface usually interacts with a constantly fluctuating mix of reactants, products, 'spectators' that do not participate in the reaction, and species that either promote or inhibit the activity of the catalyst. How molecules adsorb and dissociate under such dynamic conditions is often poorly understood. For example, the dissociative adsorption of the diatomic molecule H_2 —a central step in many industrially important catalytic processes—is generally assumed¹ to require at least two adjacent and empty atomic adsorption sites (or vacancies). The creation of active sites for H_2 dissociation will thus involve the formation of individual vacancies and their subsequent diffusion and aggregation^{2–6}, with the coupling between these events determining the activity of the catalyst surface. But even though active sites are the central component of most reaction models, the processes controlling their formation, and hence the activity of a catalyst surface, have never been captured experimentally. Here we report scanning tunnelling microscopy observations of the transient formation of active sites for the dissociative adsorption of H_2 molecules on a palladium (111) surface. We find, contrary to conventional thinking¹, that two-vacancy sites seem inactive, and that aggregates of three or more hydrogen vacancies are required for efficient H_2 dissociation.

The experiments were performed in an ultrahigh-vacuum (UHV) chamber with a total background pressure below 2×10^{-10} torr and a hydrogen partial pressure of 5×10^{-11} torr, as determined by mass spectrometry. The variable-temperature scanning tunnelling microscope (STM) apparatus⁷ was modified to cool the tip to the same temperature as the sample. The Pd(111) sample was prepared by cycles of noble-gas bombardment at 1,000 K with subsequent 'flashing' to 1,100 K. Hydrogen was introduced via a leak valve.

The work presented here concerns the final stages of H_2 adsorption and dissociation on the Pd(111) surface, when the H coverage is close to one monolayer (ML). In these conditions, the formation and dynamics of active sites can be observed most clearly. On the clean surface, hydrogen molecules dissociate readily at 37 K, the lowest temperature of the present experiments. In agreement with earlier low-energy diffraction studies⁸ and total-energy calculations^{9,10}, we have found by STM¹¹ that H atoms adsorb on the three-fold face-centred cubic (f.c.c.) sites. The H atoms appear as 15 pm (1 picometre = 10^{-12} m) depressions in the STM images, owing to reduced tunnelling probability (Fig. 1a). As the H coverage increases, an ordered structure with $\sqrt{3} \times \sqrt{3}$ R30° periodicity relative to the substrate is formed at 1/3 ML, with one occupied and two empty f.c.c. sites per unit cell. This is followed, with further H adsorption, by another $\sqrt{3} \times \sqrt{3}$ structure at 2/3 ML with two occupied and one empty f.c.c. sites per cell^{8,11}. As the empty sites in this latter structure are not contiguous, it is perhaps not surprising that the sticking coefficient for dissociative adsorption of H_2 drops substantially as the coverage approaches 2/3 ML. Above 50 K however, thermally activated diffusion produces vacancy-aggregates where H_2 molecules can dissociate, allowing the adsorption process

to continue up to the saturation value of 1 ML. The STM images of the surface close to 1 ML coverage (Fig. 1b) reveal a 1×1 structure with a small, 5-pm corrugation (in contrast with the 15 pm of the isolated H atoms), with almost all the f.c.c. sites occupied by H atoms. The surface contains defects, in the form of H vacancies (empty f.c.c. sites), which appear as bright protrusions in the images. The apparent height of these protrusions is 50 pm. Because of the 10-fold contrast difference the vacancy protrusions dominate the image contrast, and the 1×1 H areas appear rather dark (Fig. 1b).

The dynamics of vacancy diffusion and their annihilation by dissociative adsorption of H_2 at a temperature of 65 K was followed by acquiring successive STM images of the same area at 75 s per

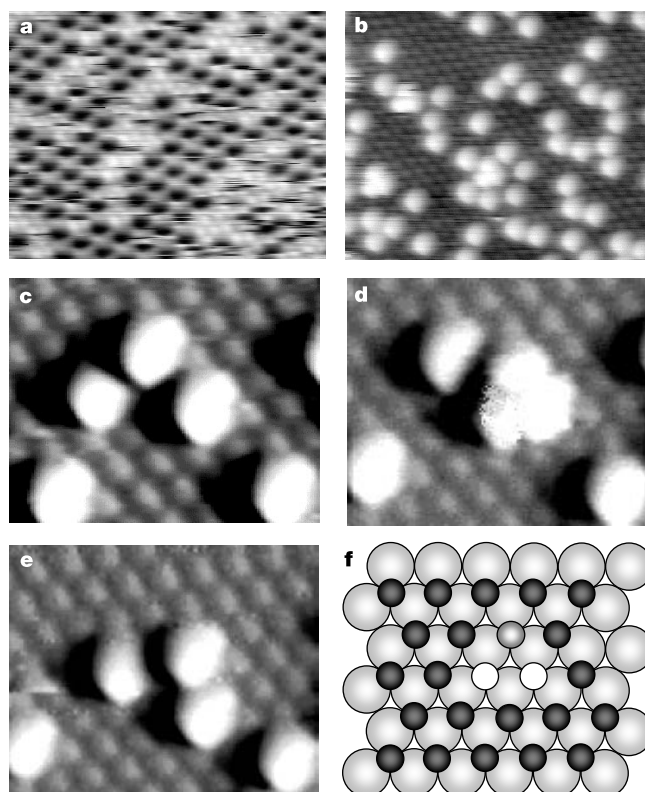


Figure 1 STM images of hydrogen atoms on Pd(111). **a**, 6.2×5 nm at a coverage of ~ 0.17 ML. The H atoms appear as dark spots (15-pm depressions) owing to reduced tunnelling current. They form islands with $\sqrt{3} \times \sqrt{3}$ R30° periodicity. The bright spots between islands correspond to Pd atoms, and have a corrugation of ~ 2 pm. **b**, 6.4×5.2 nm images acquired at a H coverage of almost 1 ML. The H atoms form an ordered 1×1 structure with weak, 5-pm corrugation. The large bright spots correspond to H vacancies, with an apparent height of about 50 pm. **c–e**, 2.0×1.5 nm images, which are frames of a movie showing the formation and dissociation of a H-vacancy pair in the almost H-saturated Pd(111) surface. To enhance contrast, these images are shown shadowed from the right. In **c**, five H vacancies are present. A few minutes later (**d**), two vacancies have moved to nearest-neighbour positions forming a vacancy pair, which is imaged as a three-lobed triangle. In **e** the vacancies have separated, and five isolated vacancies can again be seen. The drawing in **f** illustrates the mechanism explaining the triangular appearance of the pair: large circles represent Pd atoms; small circles, H atoms. The light-shaded H atom can rapidly exchange with the vacancies (white circles) by hopping over bridge sites that offer the smallest barrier to diffusion. Exchange with the H atoms surrounding the triangle requires hopping near top Pd sites and/or close to other H atoms, a process with a higher barrier and therefore much slower. As the exchange is fast compared to the STM line-scan rate, each of the three triangle sites is vacant 2/3 of the time and appears equally bright. Movies of Monte Carlo simulations of vacancy diffusion and aggregation can be viewed at <http://stm.lbl.gov> (see also Supplementary Information).

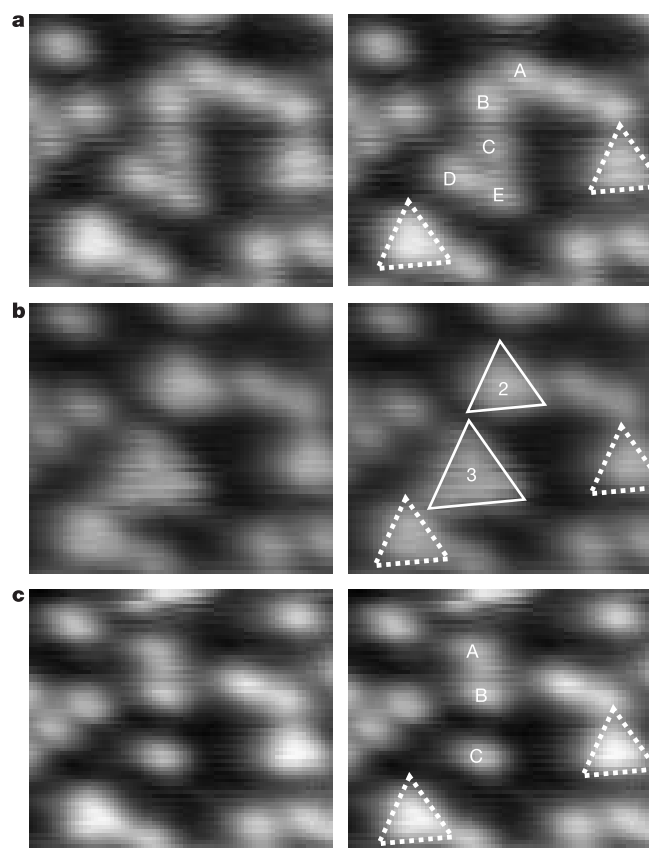


Figure 2 STM images from a movie showing the formation, separation and annihilation of H-vacancy clusters. The contrast is such that only the H vacancies are visible in the form of bright spots. The images on the left (3×2.5 nm) are repeated on the right with annotations. **a**, Five vacancies near the centre are labelled (A–E) and two triangular vacancy pairs (2V) are marked with dashed triangles for reference. In **b**, vacancies A and B have formed a 2V cluster indicated by the triangle containing the number ‘2’. Vacancies C–E have formed a six-site three-vacancy (3V) cluster, indicated by the larger triangle containing the number ‘3’. Rapid motion of the 3 H atoms inside the triangle explains the triangular appearance of these clusters. In **c** the 2V pair has separated into isolated vacancies A and B, while the 3V cluster has been annihilated by dissociative adsorption of an H_2 molecule, leaving a single remaining vacancy C. The other 2V clusters separated a few frames later.

frame. The gas-phase H_2 pressure was increased to 2×10^{-7} torr. At 65 K the observed vacancy hopping frequency was $4 \times 10^{-4} s^{-1}$, corresponding to an activation energy of ~ 0.2 eV. For comparison, the activation energy for isolated H-atom diffusion was found to be ~ 0.09 eV (ref. 11). Three frames from this experiment are shown (Fig. 1c–e). These images are shown with simulated ‘shadows’ from a light source on the right-hand side to enhance the contrast of the 1×1 H regions. In Fig. 1c, five vacancies are visible. In the next frame (Fig. 1d) two of them, near the centre, have aggregated to form a vacancy pair, which later separates (Fig. 1e). Unexpectedly, the vacancy pair has the appearance of a bright triangle occupying three nearest-neighbour f.c.c. sites. When the pair separates, two isolated vacancies are again imaged.

The triangular shape of the vacancy pair can be explained with the help of the schematic diagram in Fig. 1f. According to total-energy calculations, the lowest-energy diffusion path for H is from an f.c.c. to an adjacent hexagonal close packed (h.c.p.) site across a bridge site, then across a second bridge to a neighbouring f.c.c.⁹. The light-shaded H atom in Fig. 1f can follow this path to either vacant site without hindrance, while all other surrounding H atoms (dark shaded in the diagram) must pass close to a neighbouring H atom

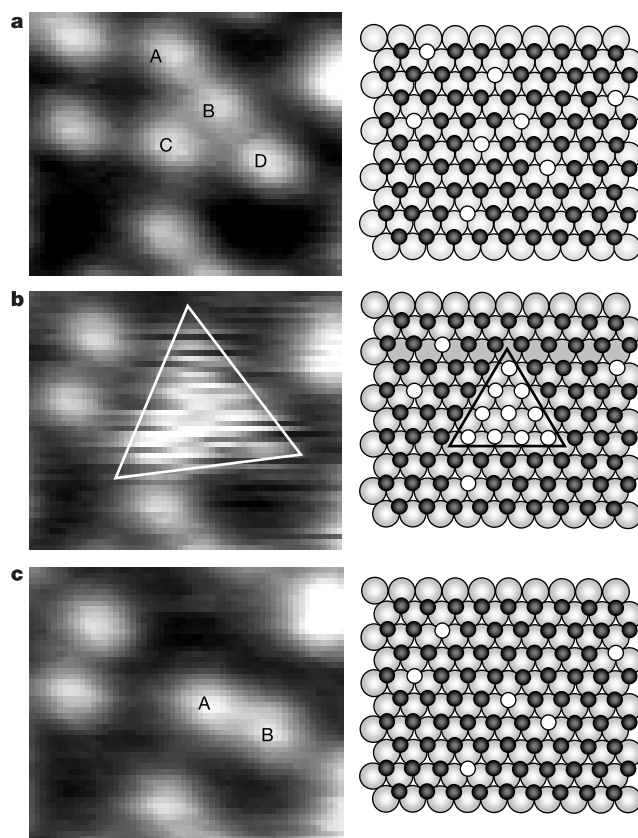


Figure 3 STM images from a movie showing the formation and annihilation of a four-vacancy (4V) cluster. The contrast is such that only the H vacancies are visible in the form of bright spots. The corresponding schematic diagrams show the H atoms and vacancies relative to the Pd(111) substrate. Letters A–D mark four vacancies near the centre in **a**. In **b**, these vacancies have formed a ten-site 4V cluster, indicated by the triangle, with 6 H atoms rapidly moving inside. In **c** the 4V cluster has been annihilated by adsorption of an H_2 molecule, leaving two remaining vacancies.

(within 0.57 of the lattice spacing) or over a top site, a less favourable path. If the unhindered diffusion rate is fast compared to the STM imaging time, the light-shaded H atom will repeatedly change positions between the three corners of the triangular cluster, and the STM tip will ‘see’ each of the sites empty for 2/3 of the time on average. As we shall see below this phenomenon is quite general, and occurs also with larger aggregates. Vacancy clusters always appear as partially occupied bright triangles bounded by close-packed {110} edges, with H atoms diffusing rapidly inside.

Events where more than two vacancies aggregate to form larger clusters were also observed (Fig. 2). In Fig. 2a we can see several isolated vacancies and two vacancy pairs. In Fig. 2b the five labelled vacancies of Fig. 2a have formed two new clusters: a vacancy pair (2V) and a triplet (3V). The 3V cluster occupies a six-site triangular area bounded by {110} edges, similar to the 2V case, with three H

Table 1 Behaviour of different vacancy clusters

Vacancy cluster size	2V	3V	4V	5V
Number observed	40	57	24	17
Total image frames	228	152	35	22
H_2 adsorption events	0	35	18	14
Mean adsorption time (s)	$>2 \times 10^4$	330	150	120
H_2 sticking coefficient	<0.0001	0.005	0.008	0.008
Cluster separation events	29	16	4	2
Mean separation time (s)	590	710	670	820

Active site statistics for vacancy clusters at 65 K in 2×10^{-7} torr H_2 . STM frame time, 75 s. H_2 sticking coefficient estimated by assuming an H_2 capture cross-section corresponding to the geometrical area of the Pd(111) unit cell per H vacancy. In addition to adsorption or separation, some clusters changed size by merging with additional vacancies.

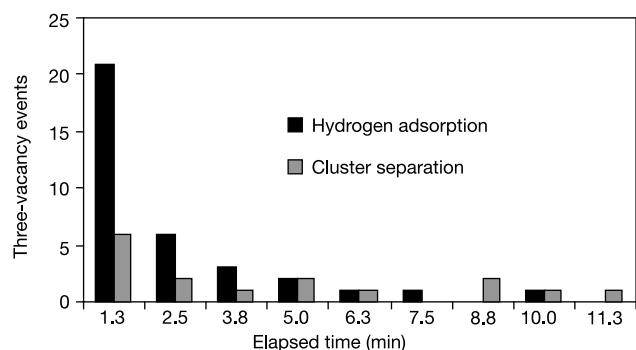


Figure 4 Lifetimes of three-vacancy clusters. Shown is the observed lifetime distribution of three-vacancy (3V) clusters at 65 K in a background pressure of $\sim 2 \times 10^{-7}$ torr H_2 . The clusters decay either by separation or by annihilation following adsorption of a H_2 molecule, which fills two of the three H vacancies.

atoms diffusing rapidly among the six sites. In contrast to the 2V clusters, the larger 3V clusters are active for H_2 dissociation (Fig. 2c). The 2V cluster has separated into isolated vacancies (A, B), while the 3V cluster has been annihilated by a dissociating H_2 molecule, leaving one vacancy (C) behind. Two other 2V clusters have remained intact during this process, but separate several frames later (not shown). The annihilation of 3V clusters occurs only in the presence of H_2 background gas.

As a final example, Fig. 3 shows the formation and annihilation of a 4V cluster. Again, this cluster appears as a triangle bounded by {110} edges and occupying an area containing ten Pd atoms. The six H atoms in the triangle diffuse rapidly among the ten sites. This cluster decayed via H_2 adsorption, leaving a pair of isolated vacancies. For this and for larger vacancy clusters, annihilation by H_2 adsorption was the dominant decay mechanism.

Figure 4 shows the observed lifetime distribution for 3V clusters in 2×10^{-7} torr of H_2 , and Table 1 gives statistics for 2V to 5V clusters. Hydrogen adsorption was frequently observed for 3V and larger clusters, always reducing the number of vacancies by two. The H_2 sticking probability can be estimated at 0.5–0.8% for 3V and larger clusters. In contrast, we never observed H_2 adsorption for a 2V cluster even though two atomic hydrogen sites are available. This finding is contrary to accepted models of diatomic molecule dissociation, and calls for a theoretical effort in modelling the adsorption step in catalysis that includes the precise atomic structure of the target sites and the role of adsorbate diffusivity in generating such sites.

As a first step to understanding the dynamics of vacancy aggregation, we performed Monte Carlo simulations of vacancy diffusion. The model, which uses only experimentally estimated values of the diffusion barriers of H atoms and vacancies (ref. 11; see also <http://stm.lbl.gov>), correctly reproduced the observed stability and the triangular appearance of the 2V, 3V and 4V clusters, and could thus help to improve our molecular-level understanding of the formation of the active sites that determine the activity of catalytically active surfaces. □

Received 13 January; accepted 4 March 2003; doi:10.1038/nature01557.

- Conrad, H., Ertl, G. & Latta, E. E. Adsorption of hydrogen on palladium single crystal surfaces. *Surf. Sci.* **41**, 435–446 (1974).
- Taylor, H. S. Theory of the catalytic surface. *Proc. R. Soc. Lond. A* **108**, 105–111 (1925).
- Boudart, M. Four decades of active centers. *Am. Sci.* **57**, 97–111 (1969).
- Ponec, V. & Sachtler, W. M. H. The reactions between cyclopentane and deuterium on nickel and nickel-copper alloys. *J. Catal.* **24**, 250–261 (1972).
- Sinfelt, J. H. *Bimetallic Catalysts: Discoveries, Concepts and Applications* (Wiley and Sons, New York, 1983).
- Somorjai, G. A. *Introduction to Surface Chemistry and Catalysis* (Wiley and Sons, New York, 1994).
- Behler, S. et al. A scanning tunneling microscope with continuous flow cryostat sample cooling. *Rev. Sci. Instrum.* **68**, 2479–2485 (1997).
- Felter, T. E., Sowa, E. C. & Van Hove, M. A. Location of hydrogen on palladium (111) studied by low-energy electron diffraction. *Phys. Rev. B* **40**, 891–899 (1989).

- Paul, J. F. & Sautet, P. Density-functional periodic study of the adsorption of hydrogen on a palladium (111) surface. *Phys. Rev. B* **53**, 8015–8027 (1996).
- Lovvik, O. M. & Olsen, R. A. Adsorption energies and ordered structures of hydrogen on Pd(111) from density functional periodic calculations. *Phys. Rev. B* **58**, 10890–10898 (1998).
- Mitsui, T., Rose, M. K., Fomin, E., Ogletree, D. F. & Salmeron, M. Hydrogen adsorption and diffusion on Pd(111). *Surf. Sci.* (submitted).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements This work was supported by the Office of Basic Energy Science of the US Department of Energy.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.S. (e-mail: salmeron@stm.lbl.gov).

Global anisotropy and the thickness of continents

Yuancheng Gung, Mark Panning & Barbara Romanowicz

Berkeley Seismological Laboratory and Department of Earth and Planetary Science, Berkeley, California 94720, USA

For decades there has been a vigorous debate about the depth extent of continental roots^{1,2}. The analysis of heat-flow³, mantle-xenolith⁴ and electrical-conductivity⁵ data all indicate that the coherent, conductive part of continental roots (the 'tectosphere') is at most 200–250 km thick. Some global seismic tomographic models agree with this estimate, but others suggest that a much thicker zone of high velocities lies beneath continental shields^{6–9}, reaching a depth of at least 400 km. Here we show that this disagreement can be reconciled by taking into account seismic anisotropy. We show that significant radial anisotropy, with horizontally polarized shear waves travelling faster than those that are vertically polarized, is present under most cratons in the depth range 250–400 km—similar to that found under ocean basins^{9,10} at shallower depths of 80–250 km. We propose that, in both cases, the anisotropy is related to shear in a low-viscosity asthenospheric channel, located at different depths under continents and oceans. The seismically defined 'tectosphere' is then at most 200–250 km thick under old continents. The 'Lehmann discontinuity', observed mostly under continents at about 200–250 km, and the 'Gutenberg discontinuity', observed under oceans at depths of about 60–80 km, may both be associated with the bottom of the lithosphere, marking a transition to flow-induced asthenospheric anisotropy.

The maximum thickness of the lithosphere, defined as a region of distinctly faster than average seismic velocities (1.5%–2%) in global S-wave velocity V_S tomographic models, ranges from 200 to 400 km, depending on the model^{6–9}. This is clear from the drop in correlation between some models from around 0.80 at 100 km to less than 0.45 at 300 km depth (Fig. 1a), which casts some doubt on the ability of global tomography to accurately resolve upper mantle structure. However, although global V_S models differ from each other significantly in the depth range 200–400 km under the main continental shields, these differences are consistent when they are classified into three categories, depending on the type of data used to derive them: 'SV' (mostly vertical or longitudinal component data, dominated by Rayleigh waves in the upper mantle), 'SH' (mostly transverse component data, dominated by Love waves), and 'hybrid' (three-component data). SH and hybrid models are better

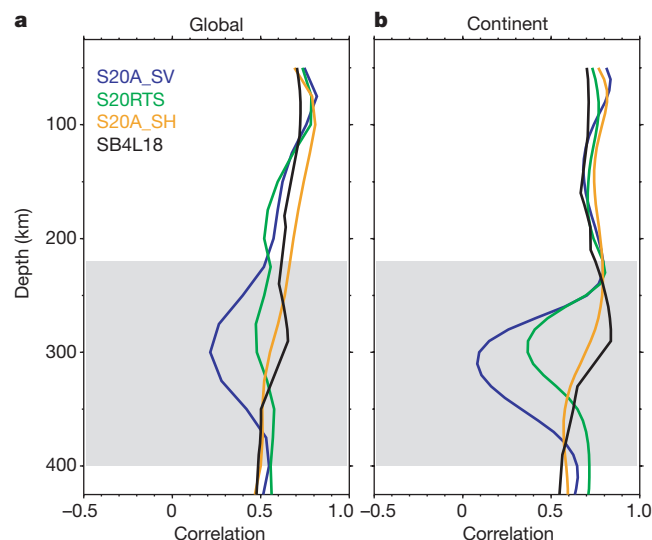


Figure 1 Correlation coefficient as a function of depth between model SAW24B16⁸, an SH model, and other global tomographic S velocity models. **a**, Correlation computed over the whole globe; **b**, correlation computed over continental areas only. Here continents include all areas of elevation greater than -0.5 km. Note that models S20A_SH⁹ (an SH model) and SB4L18⁶ (a hybrid model) correlate better with SAW24B16 than models S20A_SV⁹ and S20RTS⁷, which are both SV models. The reduced correlation in the depth range 250–400 km between SH/hybrid models and SV models is strongly accentuated over continents.

correlated with each other than with SV models. This difference is accentuated when the correlation is computed only across continental areas (Fig. 1b). Also, SH (and hybrid) models exhibit continental roots that exceed those of SV models by 100 km or more, as illustrated in Fig. 2 (see also Supplementary Fig. 1).

On the other hand, global tomographic studies that account for seismic anisotropy, either by inverting separately for V_{SV} and V_{SH} ⁹, or in the framework of more general anisotropic theory¹⁰, have

documented significant lateral variations in the anisotropic parameter $\xi = (V_{SH}/V_{SV})^2$ on the global scale. Until now, attention has mostly focused on the strong positive $\delta \ln \xi$ ($\delta \ln V_{SH} > \delta \ln V_{SV}$) observed in the central part of the Pacific Ocean in the depth range 80–200 km. The presence of this anisotropy has been related to shear flow in the asthenosphere, with a significant horizontal component. Deeper anisotropy was suggested, but not well resolved in these studies, either because the data set was limited to fundamental mode surface waves¹⁰, or because of the use of inaccurate depth sensitivity kernels⁹. In particular, it is important to verify that any differences in V_{SV} and V_{SH} observed below a depth of 200 km are not an artefact of simplified theoretical assumptions, which ignore the influence of radial anisotropy on depth sensitivity kernels (see Supplementary Fig. 2 for a comparison of isotropic and anisotropic V_S kernels).

We have developed an inversion procedure for transverse isotropy using three-component surface and body waveform data, in the framework of normal mode asymptotic coupling theory¹¹, which in particular, involves the use of two-dimensional broadband anisotropic sensitivity kernels appropriate for higher modes and body waves (see Methods section). Figure 3 shows the distributions of $\delta \ln \xi$ in the resulting degree-16 anisotropic model (SAW16AN) (for the corresponding distributions in V_{SH} , V_{SV} , see Supplementary Fig. 3), at depths of 175 km, 300 km and 400 km. At 175 km depth, the global distribution of $\delta \ln \xi$ confirms features found in previous studies, and is dominated by the striking positive $\delta \ln \xi$ ($V_{SH} > V_{SV}$) anomaly in the central Pacific^{9,10} and a similar one in the Indian Ocean. However, at depths greater than 250 km, the character of the distribution changes: positive $\delta \ln \xi$ emerges under the Canadian Shield, Siberian Platform, Baltic Shield, southern Africa, Amazonian and Australian cratons, while the positive $\delta \ln \xi$ fades out under the Pacific and Indian oceans. At 300 km depth, the roots of most cratons are characterized by positive $\delta \ln \xi$, which extend down to about 400 km. These features are emphasized in depth cross-sections across major continental shields (Fig. 4), where we compare V_{SH} and V_{SV} distributions, consistently showing deeper continental roots in V_{SH} . The presence of anisotropy at depths over 200 km, with $V_{SH} > V_{SV}$, is also consistent with some

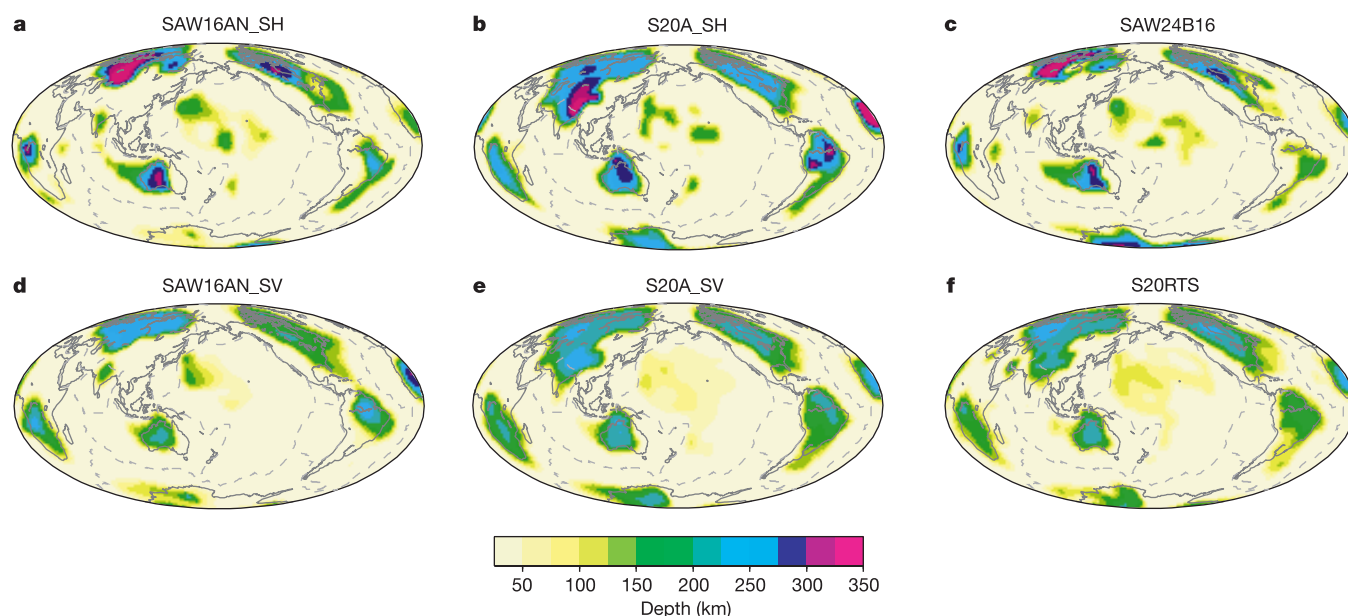


Figure 2 Maximum depth for which the velocity anomaly with respect to the reference model PREM²⁸ is greater than 2%, for different S velocity models. **a–c**, SH-type models; **d–f**, SV-type models. **c, f**, V_{SH} model SAW24B16⁸ compared to V_{SV} model S20RTS⁷. **b, e**, SH and SV parts of model S20A⁹, obtained by inverting T-component data and

Z_L-component data separately. **a, d**, SH and SV parts of anisotropic model SAW16AN discussed here. While the roots of continents generally extend to depths greater than 300–350 km in SH models, they do not exceed 200–250 km in SV models.

regional studies^{12,13}. Interestingly, the East Pacific Rise has a signature with $\delta \ln \xi < 0$ down to 300 km, indicative of a significant component of vertical flow. At 400 km depth, we also note the negative $\delta \ln \xi$ around the Pacific ring, consistent with quasi-vertical flow in the subduction zone regions in the western Pacific and South America.

There has been a long-lasting controversy regarding the interpretation of shear wave splitting observations under continents, with some authors advocating frozen anisotropy in the lithosphere¹⁴, and others, flow-induced anisotropy related to present-day plate motions¹⁵. SKS splitting measurements do not have adequate depth resolution, and inferences that have been made on the basis of a lithospheric thickness of 400 km or more under cratons need to be revisited.

Temperatures in the 250–400 km depth range exceed 1,000 °C, and are therefore too high to allow sustained frozen anisotropy in a mechanically coherent lithospheric lid on geologically relevant timescales¹⁵. Therefore we infer that the $V_{SH} > V_{SV}$ anisotropy we describe here must be related to present-day flow-induced shear, with a significant horizontal component. Such an interpretation is also compatible with results from shear wave splitting, which document the presence of anisotropy below cratons indicating

simple-shear deformation parallel to present-day plate motion, at least in North America^{16,17} and Australia¹³; some recent studies indicate that there may be two zones of SKS anisotropy under continental shields, one shallower, reflecting past geological events, and one deeper, related to present-day flow^{16,18}.

We note the similarity of the character of $V_{SH} > V_{SV}$ anisotropy, in the depth range 200–400 km under cratons, and 80–200 km under ocean basins, and we suggest that both are related to shear in the asthenosphere, the difference in depth simply reflecting the varying depth of the asthenospheric channel. Although our inference is indirect, it reconciles tomographic studies with other geophysical observations of lithospheric thickness based on heat flow³, xenoliths⁴ and mantle electrical conductivity⁵. It is also in agreement with lateral variations in attenuation on the global scale¹⁹.

Another contentious issue is the nature of the Lehmann discontinuity, and in particular the puzzling observation that it is not a consistent global feature²⁰, but is observed primarily in stable continental areas and not under oceans^{21,22}. Leven *et al.*²³ first proposed that the Lehmann might be a discontinuity in anisotropic structure, and more recent studies have suggested that it is a rheological boundary marking a transition from anisotropic to isotropic structure^{24–25}. Because the $V_{SH} > V_{SV}$ anisotropy under continental cratons is found deeper than 200 km, we propose that the Lehmann discontinuity actually marks the top of the asthenospheric layer, a transition from weak anisotropic lowermost continental lithosphere to anisotropic asthenosphere, in agreement with the inference of ref. 23. Under oceans, the lithosphere is much thinner, and the lithosphere/asthenosphere boundary occurs at much shallower depths. There is no consistently observed discontinuity around 200–250 km depth²⁰. On the other hand, the shallower Gutenberg discontinuity is often reported under oceans

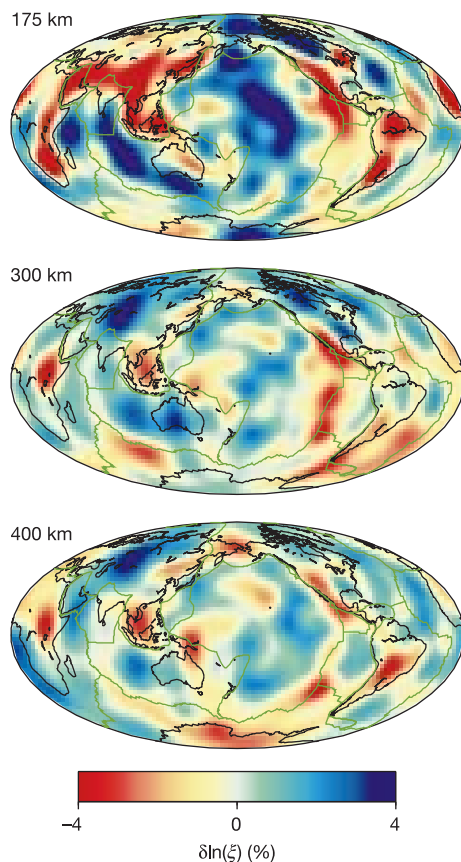


Figure 3 Maps of relative lateral variations in the anisotropic parameter $\delta \ln \xi$ of model SAW16AN at three depths in the upper mantle. $\delta \ln \xi > 0$ in regions where $V_{SH} > V_{SV}$ and $\delta \ln \xi < 0$ in regions where $V_{SV} < V_{SH}$. Lateral variations are referred to reference model PREM²⁸, which is isotropic below 220 km depth, but has significant $\delta \ln \xi > 0$ at 175 km depth. Note how the regions of strong positive $\delta \ln \xi$ shift from the central Pacific to continental areas between 175 and 300 km depth. At depths shallower than 200 km, continental shields have mostly $\delta \ln \xi < 0$, as noted previously²⁹. At 300 km depth, continental shields are no longer prominent in V_{SV} . At depths greater than 350 km, the subduction zones are more prominent in V_{SV} than in V_{SH} , resulting in $\delta \ln \xi < 0$ around the Pacific, indicative of vertical flow. The East Pacific Rise appears as a zone of vertical flow to depths in excess of 300 km. Depth resolution is of the order of ± 50 km.

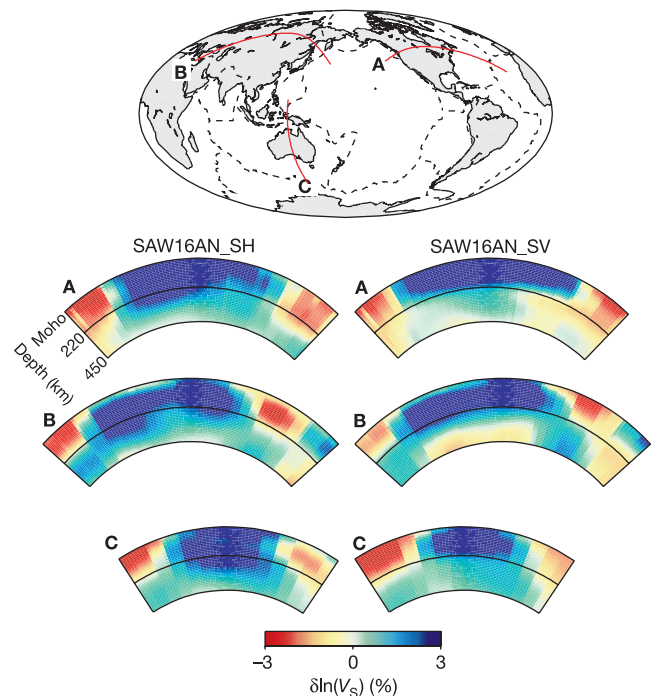


Figure 4 Depth cross-sections through three continents (see locations at top) showing the SH (left) and SV (right) components of anisotropic model SAW16AN. The SH sections consistently indicate fast velocities extending to depths in excess of 220 km, whereas the SV sections do not. In section B, the higher velocity associated with the subduction under the area north of Japan is clearly visible in SV but not so much in SH. This anisotropy may contribute to explaining why subduction zones are generally less visible in S tomographic models (mostly of the hybrid type, which thus more sensitive to SH) than in P models.

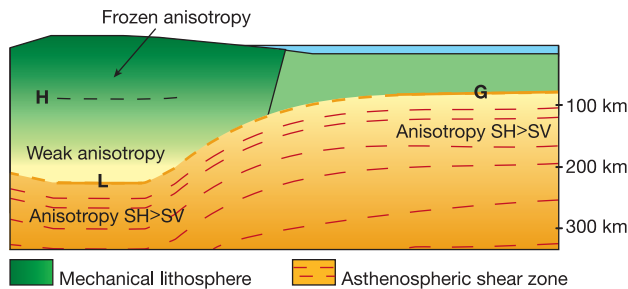


Figure 5 Sketch illustrating our interpretation of the observed anisotropy in relation to lithospheric thickness, and its relationship to the Lehmann (L) and Gutenberg (G) discontinuities. The Hales discontinuity (H), which is also shown, is generally observed as a positive impedance embedded within the continental lithosphere in the depth range 60–80 km (ref. 30). H and G may not be related.

and appears as a negative impedance reflector in studies of precursors to multiple ScS reverberations²². The difference in depth of the observed $\delta \ln \xi > 0$ anisotropy between continents and oceans is consistent with an interpretation of both the Lehmann and Gutenberg discontinuities as marking the bottom of the mechanically coherent lithosphere, in areas where it is quasi-horizontal (Fig. 5).

In this study, we consider only radial anisotropy, which in particular does not account for intermediate orientation of the fast axis of anisotropy¹⁰. We can only infer that regions with significant $\delta \ln \xi > 0$ are regions where anisotropy has a significant horizontal component, and expresses the alignment of olivine crystals in predominantly horizontal flow²⁶. In regions of transition between cratons and younger continental provinces, or between ocean and continent, the asthenospheric flow would follow the inclined shape of the bottom of the lithosphere and be less clearly detected with our approach.

Thus, the inspection of radial anisotropy in the depth range 200–400 km allows us to infer that continental roots do not extend much beyond a depth of 250 km, in agreement with other geophysical observations. The part of the mantle under old continents that translates coherently with plate motions need not be thicker than 200–250 km. Tomographic models reveal the varying depth of the top of the anisotropic asthenospheric channel, marked by the Lehmann discontinuity under continents (about 200–250 km depth), and the Gutenberg discontinuity under oceans (about 60–80 km depth). Seemingly incompatible tomographic models obtained by different researchers can thus also be reconciled: the relatively poor correlation between different models in the depth range 250–400 km is not due to a lack of resolution of the tomographic approach, but rather to the different sensitivity to anisotropy of different types of data. □

Methods

Broadband sensitivity kernels

In this study, we invert three-component long-period seismograms in the time domain (down to periods of 60 s for surface waves, and 32 s for body waves) in the framework of nonlinear asymptotic coupling theory (NACT¹¹), a normal-mode perturbation-based approach which takes into account the concentrated sensitivity of body waves to structure along the ray path, in contrast to standard approaches which assume one-dimensional kernels, an approximation which is valid only for fundamental mode surface waves. Our technique involves dividing the seismogram into wavepackets that may contain one or more seismic phases, and applying weighting factors to equalize the contribution of large and small amplitude wavepackets in the least-squares inversion.

Transverse isotropy

A transversely isotropic medium with vertical axis of symmetry is described by density ρ and five elastic parameters, usually $A = \rho V_{PH}^2$, $C = \rho V_{PV}^2$, $L = \rho V_{SV}^2$, $N = \rho V_{SH}^2$ and F . We start by considering, equivalently, the six parameters V_{SH} , V_{SV} , $\eta = F/(A - 2L)$, $V_{P_{iso}}$ (isotropic V_P), $\phi = C/A$ and ρ , with appropriate kernels for weak transverse anisotropy. To reduce the number of parameters in the inversion and keep only those that are best-

resolved ($V_{SH} = (N/\rho)^{1/2}$, $V_{SV} = (L/\rho)^{1/2}$), we assume the following scaling relations, as inferred from laboratory experiments for depths relevant to our study (that is, less than 500 km)²⁷: $\delta \ln V_{P_{iso}} = 0.5 \delta \ln V_{S_{iso}}$, $\delta \ln \eta = -2.5 \delta \ln \xi$ and $\delta \ln \phi = -1.5 \delta \ln \xi$, with $\delta \ln V_{S_{iso}} = 2/3 \delta \ln V_{SV} + 1/3 \delta \ln V_{SH}$ (under the assumption of weak anisotropy; $V_{S_{iso}}$ is isotropic V_S and $\delta \ln \rho = 0.3 \delta \ln V_{S_{iso}}$). We have verified that the main features in our results are not affected by the particular values chosen in these relations. Starting from our most recent tomographic models, SAW24B16⁸ for V_{SH} and SAW16B19¹⁹ for V_{SV} , we invert for perturbations in V_{SH} and V_{SV} in a spherical harmonics expansion up to degree 16 laterally. Vertical parametrization is in terms of cubic splines. Since our sampling of the lowermost mantle with SV-sensitive body waves is limited, in order to avoid bias from anisotropy in D' , we have restricted our inversion to the top 1,500 km of the mantle, and chosen the body waveforms to include in the data set accordingly.

We have checked that our results, and in particular the observation of radial anisotropy under continents at depths greater than 200 km is not the result of artefacts due to poor resolution in the inversion for either V_{SH} or V_{SV} , by performing synthetic tests. For example, Supplementary Fig. 4 shows the results of an experiment in which synthetic transverse component seismograms have been computed for a starting SV model (no roots below 250 km), mimicking the actual distribution of our data set, and then reinverted for an SH model. No deep continental roots are apparent in the resulting final model.

Assuming lattice preferred orientation of anisotropic minerals such as olivine, and as illustrated, for example, in ref. 26, a large-scale predominantly horizontal flow is characterized by a positive $\delta \ln \xi$ and also by significant SKS wave splitting. The direction of the fast axis inferred from the latter is related to the direction of the flow in the horizontal plane. Coupling between Love and Rayleigh waves that may arise in the case of anisotropy with a non-vertical axis of symmetry affects mainly the wave amplitudes. Because we are primarily fitting the phase part of the seismograms, such coupling should have little effect on our results.

Received 18 September 2002; accepted 4 March 2003; doi:10.1038/nature01559.

- Jordan, T. H. The continental lithosphere. *Rev. Geophys. Space Phys.* **13**, 1–12 (1975).
- Anderson, D. L. The deep structure of continents. *J. Geophys. Res.* **84**, 7555–7560 (1990).
- Jaupart, C., Mareschal, J. C. & Guillou-Frotier, L. Heat flow and thickness of the lithosphere in the Canadian Shield. *J. Geophys. Res.* **103**, 15269–15286 (1998).
- Rudnick, R., McDonough, W. & O'Connell, R. Thermal structure, thickness and composition of continental lithosphere. *Chem. Geol.* **145**, 395–411 (1998).
- Hirth, G., Evans, R. L. & Chave, A. D. Comparison of continental and oceanic mantle electrical conductivity: Is the Archean lithosphere dry? *Geochim. Geophys. Geosyst.* **1**, 2000GC000048 (2000).
- Masters, G., Johnson, S., Laske, G. & Bolton, B. A shear-velocity model of the mantle. *Phil. Trans. R. Soc. Lond. A* **354**, 1385–1411 (1996).
- Ritsema, J., van Heijst, H. & Woodhouse, J. H. Complex shear wave velocity structure imaged beneath Africa and Iceland. *Science* **286**, 1925–1928 (1999).
- Méginn, C. & Romanowicz, B. The 3D shear velocity structure of the mantle from the inversion of body, surface and higher mode waveforms. *Geophys. J. Int.* **143**, 709–728 (2000).
- Ekström, G. & Dziewonski, A. M. The unique anisotropy of the Pacific upper mantle. *Nature* **394**, 168–172 (1998).
- Montagner, J. P. What can seismology tell us about mantle convection? *Rev. Geophys.* **32**(2), 135–137 (1994).
- Li, X. D. & Romanowicz, B. Comparison of global waveform inversions with and without considering cross branch coupling. *Geophys. J. Int.* **121**, 695–709 (1995).
- Tong, C., Gudmundsson, O. & Kennett, B. L. N. Shear wave splitting in refracted waves returned from the upper mantle transition zone beneath northern Australia. *J. Geophys. Res.* **99**, 15783–15797 (1994).
- Debayle, E. & Kennett, B. L. N. Anisotropy in the Australasian upper mantle from Love and Rayleigh waveform inversion. *Earth Planet. Sci. Lett.* **184**, 339–351 (2000).
- Silver, P. G. Seismic anisotropy beneath the continents: probing the depths of geology. *Annu. Rev. Earth Planet. Sci.* **24**, 385–432 (1996).
- Vinnik, L. P., Makeyeva, L. I., Milev, A. & Usenko, Y. Global patterns of azimuthal anisotropy and deformations in the continental mantle. *Geophys. J. Int.* **111**, 433–447 (1992).
- Fouch, M. J., Fischer, K. M., Parmentier, E. M., Wyssession, M. E. & Clarke, T. J. Shear wave splitting, continental keels, and patterns of mantle flow. *J. Geophys. Res.* **105**, 6255–6275 (2000).
- Bokelmann, G. H. R. Which forces drive North America? *Geology* **30**(11), 1027–1030 (2002).
- Levin, V., Menke, W. & Park, J. Shear wave splitting in the Appalachians and the Urals: a case for multilayered anisotropy. *J. Geophys. Res.* **104**, 17975–17993 (1999).
- Romanowicz, B. & Gung, Y. Superplumes from the core-mantle boundary to the lithosphere: implications for heat flux. *Science* **296**, 513–516 (2002).
- Shearer, P. Seismic imaging of upper mantle structure with new evidence for a 520 km discontinuity. *Nature* **344**, 121–126 (1990).
- Gu, Y. M., Dziewonski, A. M. & Ekström, G. Preferential detection of the Lehmann discontinuity beneath continents. *Geophys. Res. Lett.* **28**, 4655–4658 (2001).
- Revenaugh, J. & Jordan, T. H. Mantle layering from ScS reverberations. 3. The upper mantle. *J. Geophys. Res.* **96**, 19781–19810 (1991).
- Leven, J. H., Jackson, I. & Ringwood, A. E. Upper mantle seismic anisotropy and lithospheric decoupling. *Nature* **289**, 234–239 (1981).
- Karato, S. I. On the Lehmann discontinuity. *Geophys. Res. Lett.* **19**, 2255–2258 (1992).
- Gaherty, J. B. & Jordan, T. H. Lehmann discontinuity as the base of an anisotropic layer beneath continents. *Science* **268**, 1468–1471 (1995).
- Montagner, J.-P. Upper mantle low anisotropy channels below the Pacific Plate. *Earth Planet. Sci. Lett.* **202**, 263–274 (2002).
- Montagner, J. P. & Anderson, D. L. Petrological constraints on seismic anisotropy. *Phys. Earth Planet. Int.* **54**, 82–105 (1989).
- Dziewonski, A. M. & Anderson, D. L. Preliminary reference earth model. *Phys. Earth Planet. Int.* **25**, 297–356 (1981).
- Babuska, V., Montagner, J. P., Plomerova, J. & Girardin, N. Age-dependent large-scale fabric of the mantle lithosphere as derived from surface wave velocity anisotropy. *Pure Appl. Geophys.* **121**, 257–280 (1998).

30. Levin, V. & Park, J. Shear zones in the Proterozoic lithosphere of the Arabian Shield and the nature of the Hales discontinuity. *Tectonophysics* 323(3–4), 131–148 (2000).

Supplementary Information accompanies the paper on *Nature's* website (<http://www.nature.com/nature>).

Acknowledgements We thank J. Park, B. Kennett and J. P. Montagner for constructive comments on this manuscript. This work was supported through a grant from the National Science Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to B.R. (e-mail: barbara@seismo.berkeley.edu).

Soil invertebrate fauna enhances grassland succession and diversity

Gerlinde B. De Deyn*, Ciska E. Raaijmakers*, H. Rik Zoomer†, Matty P. Berg‡, Peter C. de Ruiter‡, Herman A. Verhoef‡, T. Martijn Bezemer* & Wim H. van der Putten*

* Department of Multitrophic Interactions, Centre for Terrestrial Ecology, Netherlands Institute of Ecology (NIOO-KNAW), P.O. Box 40, 6666 ZG, Heteren, The Netherlands

† Department of Animal Ecology, Institute of Ecological Science, Vrije Universiteit De Boelelaan 1085, 1081 HV Amsterdam, The Netherlands

‡ Department of Environmental Sciences, Copernicus Research Institute for Sustainable Development and Innovation, Utrecht University, P.O. Box 80115, 3508 TC Utrecht, The Netherlands

One of the most important areas in ecology is to elucidate the factors that drive succession in ecosystems and thus influence the diversity of species in natural vegetation. Significant mechanisms in this process are known to be resource limitation^{1–3} and the effects of aboveground vertebrate herbivores^{4,5}. More recently, symbiotic and pathogenic soil microbes have been shown to exert a profound effect on the composition of vegetation^{6–9} and changes therein^{10,11}. However, the influence of invertebrate soil fauna on succession has so far received little attention^{12,13}. Here we report that invertebrate soil fauna might enhance both secondary succession and local plant species diversity. Soil fauna from a series of secondary grassland succession stages selectively suppress early successional dominant¹⁴ plant species, thereby enhancing the relative abundance of subordinate¹⁴ species and also that of species from later succession stages. Soil fauna from the mid-succession stage had the strongest effect. Our results clearly show that soil fauna strongly affects the composition of natural vegetation and we suggest that this knowledge might improve the restoration and conservation of plant species diversity.

The succession of species in natural communities is one of the most fundamental concepts in ecology¹⁵. Successional patterns in vegetation can be reasonably well predicted owing to their correlation with changes in resource availability, whereas the competitive abilities of plants for these resources determines which species might become dominant². However, biotic interactions above and below ground can greatly change the outcome of these competitive interactions. Aboveground vertebrate herbivores can indirectly benefit subdominant plant species through selective feeding on dominants^{4,5}. Below the soil surface, root symbionts can enhance plant species diversity by improving the nutrient uptake and growth

of subdominants⁶, and root pathogens can do so by suppressing dominant host plant species⁷. Selective improvement or suppression of plant species can eventually lead to cyclic^{8,9,16} or unidirectional successional¹¹ shifts in local species composition.

Although soil microorganisms seem to have a profound effect on vegetation succession and plant species diversity, the role of invertebrate soil fauna has not yet been resolved. In early secondary succession grasslands, the application of soil insecticides reduces the abundance of root-feeding insects, resulting in the promotion of early successional forbs¹², suggesting that root-feeding insects might enhance the process of succession¹³. Similar effects of root-feeding insects were confirmed in microcosm¹⁷ and mesocosm¹⁸ studies. In mid-succession extensively grazed grasslands, cyclic replacement of grass and sedge species were correlated primarily with the abundance of plant-feeding nematodes¹⁹. However, soil organisms with a larger body size seem to overrule the effects of smaller soil organisms¹⁸, casting doubt on the results of studies that focus on the role of specific components of invertebrate soil fauna.

To clarify the specific contribution of the invertebrate soil fauna (hereafter called 'soil fauna') to secondary vegetation succession and to plant species diversity, we conducted an experiment in which natural soil fauna communities (Table 1) were applied to combinations of plant species from production, restoration and conservation grasslands. We constructed experimental grassland communities in which plant species from a production grassland (early succession stage), from a grassland subjected to nature restoration management during the past 20 years (mid-succession) and from a species-rich natural grassland community (the target of secondary grassland succession) were grown in mixed stands. All plant communities were established in sterilized soil and were inoculated with soil fauna from one of the three successional stages. This made it possible to determine the role of the soil fauna from a range of secondary succession grasslands on the outcome of interactions between the plant species from their own and from the other two successional stages. The soil fauna added represented the densities and composition of the field communities and included microfauna (nematodes), mesofauna (micro-arthropods) and macrofauna (beetle larvae).

Our results show that the addition of all three successional communities of soil fauna resulted in significant shifts in the

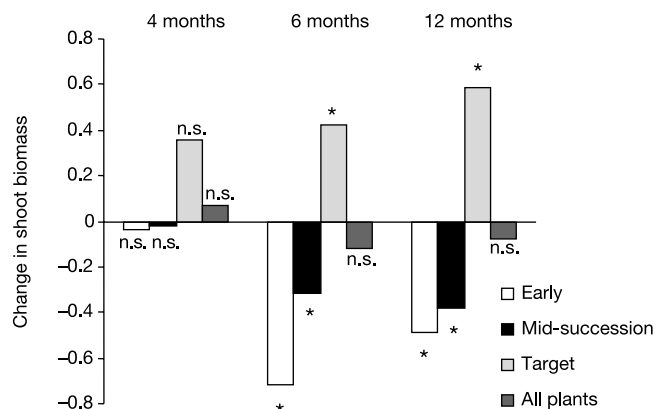


Figure 1 Change in the shoot biomass of early and mid-succession plant species and plants from the target grassland community in the presence of soil fauna relative to the control (no soil fauna added) calculated as $(B_1 - B_0)/B_0$, where B_1 and B_0 are the mean shoot biomasses with and without soil fauna added, respectively. The experimental grassland communities were grown in microcosms for 1 year. Addition of the soil fauna impaired the early and mid-succession plants and stimulated the development of plants from the target grassland community. Data presented are averages over all treatments with different origins of soil fauna. Significance: asterisk, $P < 0.05$; n.s., not significant.

Table 1 Numbers and biomass of soil fauna in the inoculum from the grasslands

Organism group	Individuals per treatment (inoculum origin)			Biomass ($\mu\text{g dry wt}$) per individual ²⁷
	Early	Mid-succession	Target	
Total nematodes	29,430 $\pm$ 9,124	28,118 $\pm$ 6,135	12,460 $\pm$ 3,466	0.33
Root-feeding nematodes	7,235 $\pm$ 3,002	9,130 $\pm$ 2,662	3,500 $\pm$ 631	0.33
Elateridae larvae	0	11 $\pm$ 7	5 $\pm$ 3	9,000
Collembola	590 $\pm$ 231	413 $\pm$ 139	713 $\pm$ 117	7.62
Mites	196 $\pm$ 59	1,698 $\pm$ 325	2,638 $\pm$ 400	1.66

Data are means $\pm$ s.e.m. ($n = 4$). Nematodes were extracted from each grassland by decantation of 100 soil cores, each 2.5 cm in diameter and 10 cm deep. Collembola, mites and Elateridae larvae were isolated from soil cores collected from the top 5-cm soil layer of each grassland; the area extracted was equal to that of the inoculated microcosms. The soil cores were extracted with Tullgren funnels²⁶ (30 $\times$ 30 cm², 30 °C top, 12 °C at the bottom). The micro-arthropods were collected from the funnels and inoculated every second day for 3 weeks.

plant communities towards the dominance of the plant species from the late succession (target) community ($P < 0.01$), whereas in the control soils the plant communities ultimately became dominated by mid-succession plant species. To identify the mechanisms underlying the enhancement of succession by the soil fauna, we examined the root and shoot biomass distributions of the various plant species in relation to the various soil communities. The soil fauna decreased the shoot biomass of the early succession plant species after 6 months ($P_6 = 0.006$) and 12 months ($P_{12} = 0.025$), as well as plant species from the mid-succession stage ($P_6 = 0.032$, $P_{12} = 0.015$), whereas the shoot biomass of the target plant species was increased ($P_6 = 0.047$, $P_{12} = 0.021$; Fig. 1). The soil fauna did not affect the total shoot biomass after 6 or 12 months but decreased the total root biomass at the end of the experiment by about one-half ($P = 0.005$). These effects on the roots indicate that the differences are due mainly to root herbivory and that the effects increase with time. The initial biomass of the root-feeding soil fauna was negatively correlated with both final root biomass and final shoot biomass ($R^2 = 0.24$, $P < 0.01$; $R^2 = 0.18$, $P < 0.05$, respectively). There was no significant correlation between collembola, or mites, and root biomass. Fungal feeders in these soil taxa might have been feeding on fungal hyphae that could develop from propagules that recolonized the microcosms throughout the course of the experiment.

Besides favouring plant species from the target site relative to plant species from early and mid-succession grasslands, addition of the soil fauna also enhanced plant species diversity. In the presence of soil fauna, the proportion of the dominant plant species in the total biomass decreased, resulting in enhanced plant evenness ($P = 0.023$; Fig. 2). Plant evenness, based on shoot biomass, was positively correlated with the initial biomass of the root-feeding soil fauna ($R^2 = 0.25$, $P < 0.01$), and negatively with the root biomass at the end of the experiment ($R^2 = 0.13$, $P < 0.05$). Although it was not possible to determine the root masses of individual plant species, our results suggest that the invertebrate root herbivores were selectively feeding on roots of dominant plants. Plant evenness was enhanced primarily by the soil fauna from the mid-succession

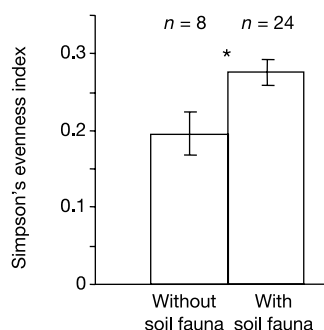


Figure 2 Simpson's evenness index (mean $\pm$ s.e.m.) of microcosms with experimental grassland communities without and with soil fauna. The soil fauna enhanced the evenness of the plant community (ANOVA; $P < 0.05$).

grassland ($P < 0.05$). In particular, the mid-succession soil fauna most selectively decreased roots and shoots of the mid-succession grasses (Fig. 3), which became the dominant plant species in controls lacking soil fauna (data for individual plant species are not shown).

Root-feeding soil fauna of the early succession communities was made up predominantly of root-feeding nematodes, whereas the faunal communities from the mid-succession and target sites consisted mainly of root-feeding nematodes and Elateridae (click beetle) larvae. Plant evenness was correlated best with initial root-feeding nematode biomass ($R^2 = 0.42$, $P = 0.0001$ for the root-feeding nematodes alone, and $R^2 = 0.25$, $P < 0.01$ for nematodes and Elateridae larvae together). The selective effects of soil fauna on the plant communities might therefore be enhanced by differences in community composition, because the nematode taxa present in the mid-succession grassland were more similar to that from the early grassland than to that from the target grassland (Sørensen's similarity index I_S ; $P < 0.001$).

In summary, our results show that the soil fauna communities from early succession, mid-succession and target stages of secondary grassland profoundly enhance vegetation succession and the homogeneity of the plant community by reducing the biomass of the dominant plant species. The strongest effect was found with the soil fauna from the mid-succession grassland, while plant species from this succession stage were affected most. Reduction of the root mass of the dominant plants by the soil fauna provided an indirect advantage for the subdominant plant species, which were only marginally suppressed in the presence of soil fauna. This was confirmed by the negative correlation between the initial biomass

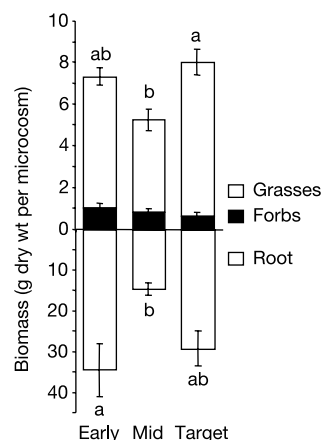


Figure 3 Effect of soil fauna from early, mid-succession and target grassland communities on total root and shoot biomasses of grasses and forbs (mean $\pm$ s.e.m.; $n = 8$). The plant communities in the microcosms were established from grassland plants from early and mid-secondary succession stages and species from the target (a semi-natural grassland). The soil fauna from the mid-succession grassland reduced the biomass of grass shoots and that of the total amount of roots (a and b denote significant differences in grass shoot or root biomass, Tukey; $P < 0.05$).

of the root feeders and the biomass of plant shoots and roots at the termination of the experiment.

The effects of soil fauna on succession are analogous to the feeding of aboveground vertebrate herbivores^{5,20} but differ markedly from the effects of arbuscular mycorrhizal fungi, which directly improve growth conditions of the subordinate plants⁶. The contribution of soil fauna and of mycorrhizal fungi to plant community processes in grasslands seems to be additive rather than counter-acting¹⁸. The net effect of the soil community on early and mid-succession and on plant species diversity seems to be due to a selective suppression of the dominant plant species by soil fauna. This might be attributed to higher root quality and accessibility (more suitable nutrient composition or less chemical or mechanical defence) or to lower tolerance to herbivory of the early succession plant species. By contrast, in late succession stages, where nutrients become limited, the growth conditions of the subordinate plant species might be selectively improved by mycorrhizal fungi¹⁰. However, their effect depends on the grassland community studied, because the competitive outcome is driven by the differential host plant responses of dominant and subordinate plant species to colonization by mycorrhizal fungi²¹. Our results show that there is a strong linkage between succession in vegetation and in soil community composition, and that the patterns observed above ground are the result of a continuous feedback process²² between plants and soil organisms. Restoration and conservation of biodiversity in natural vegetation should therefore not be focused exclusively on reducing soil fertility or on introducing aboveground vertebrate grazers. A better understanding of the dynamic contribution of soil organisms to plant community processes is also important. Including the role of soil fauna might profoundly enhance the predictability of succession and might therefore affect the success of programmes aimed at the restoration and conservation of biodiversity. □

Methods

Microcosm design maintenance and collection

We established 32 microcosms of plant species mixtures from (1) recently abandoned production grassland (early); (2) grassland under restoration for 20 years (mid-succession); and (3) species-rich natural grassland, the biodiversity restoration and conservation target (target). Experimental units consisted of microcosms (17 cm long, 17 cm wide and 22 cm deep) filled with γ -irradiated (25 kGy) sterilized soil from the mid-succession grassland, planted with 24 seedlings of *Poa trivialis* L., *Lolium perenne* L., *Stellaria media* L., *Rumex obtusifolius* L. (early species), *Agrostis capillaris* L., *Festuca rubra* L., *Plantago lanceolata* L., *Prunella vulgaris* L. (mid-succession species), *Anthoxanthum odoratum* L., *Festuca ovina* L. and *Campanula rotundifolia* L. (target species). In addition, owing to poor germination of *Succisa pratensis* L., half of the replicates received two *Centaurea jacea* L. seedlings, and the other half received one *Centaurea jacea* L. and one *Succisa pratensis* L. seedling (target species). The seeds were provided by a small specialized company and obtained from the field. One week after germination the seedlings were planted in a grid of 6 × 4 positions. Each of the eight replicates was assigned a different plant configuration to avoid positioning effects. The plants were watered daily; once a week the initial soil moisture level was measured by weighing and readjusted to 20%.

After 6 weeks the soil fauna was added (Table 1). To prevent the organisms from spreading from one microcosm to the other, a rim of Fluon was applied to the inner and outer edges of the microcosms and trolleys. Three trolleys were loaded with each microcosm with the same origin of soil fauna (early, mid-succession or target) or without inoculum (control). The trolleys with microcosms were maintained in a greenhouse in a random arrangement. There was minimally 16 h of light (240 W m⁻²) and a light/dark temperature regime of 21/16 °C. Every 2 weeks the positions of the trolleys were changed at random to minimize position effects.

After 4 and 6 months, shoots were clipped at 4 cm above the soil surface and the dry weights of all individual plant species were determined after 2 days at 75 °C. After 12 months, shoots of each plant species were clipped at ground level and total root biomass was collected, dried and weighed. Plant evenness was calculated as Simpson's evenness index²³ ($SIE = (1/\sum p_i^2)(1/S)$, where p_i is the proportional contribution of the i th species to the canopy and S is the number of species present). To determine the numbers and identity of the nematodes in the microcosms, six soil cores 2.5 cm in diameter and 20 cm deep (the entire depth of the microcosm) were collected and extracted with Oostenbrink elutriators²⁴. We analysed four replicates for nematodes; micro-arthropods were collected from the remaining four replicates. All nematodes present in 10% of the total extracted soil volume were counted and identified to genus level. The genera were distributed into feeding groups as described in ref. 25. The similarity in taxon composition was calculated as the Sørensen similarity index $I_S = 2J/(A + B)$, with J the number of taxa present in both

samples, A the number of taxa in sample A and B the number of taxa in sample B . At final collection micro-arthropods were isolated from soil cylinders (10 cm in diameter, 5 cm deep) with Tullgren funnels²⁶ (10 cm in diameter by 6 cm deep, 30 °C top, 5 °C at the bottom) and determined to species level (Collembola) or higher taxonomic levels (mites). The Collembola and mites were considered to be microbivores; the Elateridae larvae were considered to be root feeders.

Statistics

The data were analysed with Statistica version 10.0. All plant data were analysed by one-way analysis of variance (ANOVA), using soil fauna inoculum or soil fauna origin as factors. Normality was checked for by the Kolmogorov–Smirnov (Lilliefors) test, homoscedasticity with the Levene test. The statistical difference between groups was determined by Tukey's multiple-range test, with $P < 0.05$ indicating significance. If normality was not achieved, data were transformed before statistical analysis. The total shoot biomass and the shoot biomasses of early and mid-succession plant species were square-root transformed, whereas root biomasses were natural-logarithm transformed before analysis. Plant biomass per harvest was analysed as total biomass and biomass per plant successional origin and plant type (for example grasses versus forbs).

Received 5 February; accepted 6 March 2003; doi:10.1038/nature01548.

- Al-Mufti, M. M., Sydes, C. L., Furness, S. B., Grime, J. P. & Band, S. R. A quantitative analysis of shoot phenology and dominance in herbaceous vegetation. *J. Ecol.* **65**, 759–791 (1977).
- Tilman, D. *Resource Competition and Community Structure* (Princeton Univ. Press, Princeton, New Jersey, 1982).
- Marrs, R. H. Soil fertility and nature conservation in Europe: Theoretical considerations and practical management solutions. *Adv. Ecol. Res.* **24**, 241–300 (1993).
- Crawley, M. J. *Plant Ecology* (Blackwell Science, Oxford, 1997).
- Olf, H. & Ritchie, M. E. Effects of herbivores on grassland plant diversity. *Trends Ecol. Evol.* **13**, 261–265 (1998).
- van der Heijden, M. G. A. *et al.* Mycorrhizal fungal diversity determines plant biodiversity, ecosystem variability and productivity. *Nature* **396**, 69–72 (1998).
- Bever, J. D. Feedback between plants and their soil communities in an old field community. *Ecology* **75**, 1965–1977 (1994).
- Packer, A. & Clay, K. Soil pathogens and spatial patterns of seedling mortality in a temperate tree. *Nature* **404**, 278–281 (2000).
- Klironomos, J. N. Feedback with soil biota contributes to plant rarity and invasiveness in communities. *Nature* **417**, 67–70 (2002).
- Johnson, N. C., Zak, D. R., Tilman, D. & Pfleger, F. L. Dynamics of vesicular–arbuscular mycorrhizae during old field succession. *Oecologia* **86**, 349–358 (1991).
- van der Putten, W. H., Van Dijk, C. & Peters, B. A. M. Plant-specific soil-borne diseases contribute to succession in foreland vegetation. *Nature* **362**, 53–56 (1993).
- Brown, V. K. & Gange, A. C. Differential effects of above- and below-ground insect herbivory during early plant succession. *Oikos* **54**, 67–76 (1989).
- Brown, V. K. & Gange, A. C. Secondary plant succession: how is it modified by insect herbivory? *Vegetatio* **101**, 3–13 (1992).
- Grime, J. P. Benefits of plant diversity to ecosystems: immediate, filter and founder effects. *J. Ecol.* **86**, 902–910 (1998).
- Clements, F. E. *Plant Succession: An Analysis of the Development of Vegetation* (Carnegie Institution of Washington, Washington, DC, 1916).
- Blomqvist, M. M., Olff, H., Blaauw, M. B., Bongers, T. & van der Putten, W. H. Interactions between above- and belowground biota: Importance for small-scale vegetation mosaics in a grassland ecosystem. *Oikos* **90**, 582–598 (2000).
- Wardle, D. A. & Barker, G. M. Competition and herbivory in establishing grassland communities: implications for plant biomass, species diversity and soil microbial activity. *Oikos* **80**, 470–480 (1997).
- Bradford, M. A. *et al.* Impacts of soil faunal community composition on model grassland ecosystems. *Science* **298**, 615–618 (2002).
- Olff, H., Hoorens, B., de Goede, R. G. M., van der Putten, W. H. & Gleichman, J. M. Small-scale shifting mosaics of two dominant grassland species: the possible role of soil-borne pathogens. *Oecologia* **125**, 45–54 (2000).
- Carson, W. P. & Root, R. B. Herbivory and plant species coexistence: Community regulation by an outbreaking phytophagous insect. *Ecol. Monogr.* **70**, 73–99 (2000).
- Hartnett, D. C. & Wilson, G. W. T. Mycorrhizae influence plant community structure and diversity in tallgrass prairie. *Ecology* **80**, 1187–1195 (1999).
- Bever, J. D., Westover, K. M. & Antonovics, J. Incorporating the soil community into plant population dynamics: the utility of the feedback approach. *J. Ecol.* **85**, 561–573 (1997).
- Simpson, E. H. Measurement of diversity. *Nature* **163**, 688 (1949).
- Oostenbrink, M. in *Nematology* (eds Sasser, J. N. & Jenkins, W. R.) 85–102 (Univ. of North Carolina Press, Chapel Hill, North Carolina, 1960).
- Yeates, G. W., Bongers, T., de Goede, R. G. M., Freckman, D. W. & Georgieva, S. S. Feeding habits in soil nematode families and genera—an outline for soil ecologists. *J. Nematol.* **25**, 315–331 (1993).
- van Straalen, N. M. & Rijninks, P. C. The efficiency of Tullgren apparatus with respect to interpreting seasonal changes in age structure of soil arthropod populations. *Pedobiologia* **24**, 197–209 (1982).
- Petersen, H. & Luxton, M. A comparative analysis of soil fauna and their role in decomposition processes. *Oikos* **39**, 287–388 (1982).

Acknowledgements We thank J. A. Harvey for a critical reading of a previous version of the manuscript. This work was funded by the Dutch NWO-ALW Stimulation Programme Biodiversity.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to G.D.D. (e-mail: g.deedyn@nioo.knaw.nl).

Explaining the excess of rare species in natural species abundance distributions

Anne E. Magurran*† & Peter A. Henderson†‡

* Gatty Marine Laboratory, School of Biology, University of St Andrews, St Andrews, Fife KY16 8LB, UK

‡ Pisces Conservation Ltd, IRC House, The Square, Pennington, Lymington, Hampshire SO41 8GN, UK

† The authors contributed equally to this work

The observation that a few species in ecological communities are exceptionally abundant, whereas most are rare, prompted the development of species abundance models^{1–3}. Nevertheless, despite the large literature on the commonness and rarity of species inspired by these pioneering studies, some widespread empirical patterns of species abundance resist easy explanation⁴. Notable among these is the observation⁵ that in large assemblages there are more rare species than the log normal model predicts^{6,7}. Here we use a long-term (21-year) data set, from an estuarine fish community, to show how an ecological community can be separated into two components. Core species, which are persistent, abundant and biologically associated with estuarine habitats, are log normally distributed. Occasional species occur infrequently in the record, are typically low in abundance and have different habitat requirements; they follow a log series distribution. These distributions are overlaid, producing the negative skew that characterizes real data sets.

Concern about rapid biodiversity loss has intensified the need to understand community structure. It is generally accepted that most distributions of species abundance in large assemblages tend towards the log normal^{13,7,8}. However, whereas the ‘canonical hypothesis’³ was the focus of debate in the past half-century^{8,9}, it is negative skew that now captures attention. Two neutral theories predict species abundance distributions that replicate the negative skew observed in empirical data sets. The unified theory of biodiversity and biogeography⁶ develops a new species abundance distribution, the zero-sum multinomial, in which the degree of negative skewness is a function of community size and immigration

rate. The self-similarity model¹⁰ predicts that the change in species richness with area is constant across all spatial scales. Both approaches assume the ecological equivalence of all species in the community. A process of multi-dimensional niche subdivision has been postulated⁹ that also leads to negative skew⁵. We argue that the observed patterns can be more parsimoniously explained by dividing an assemblage into two components—persistent and occasional species—and without invoking neutrality.

Here we examine the relative abundance of species in an exceptionally large data set—a 21-year investigation (with monthly sampling) of a fish community at Hinkley Point in the Bristol Channel, UK, in which $S = 80$ species and $N > 96,000$ individuals were recorded (see Methods). Our data clearly show that the maximum abundance of a species in the year in which it is most abundant is a function of the number of years for which it has been recorded (Fig. 1). The commonness and rarity of species in the assemblage is thus related to their permanence. Three species (sprat (*Sprattus sprattus*), sand goby (*Pomatoschistus minutus*) and whiting (*Merlangius merlangus*)) continuously dominate and together account for an average 70% of total abundance (by weight or by number). Adding the other 28 core species brings the total to 99%. The 49 infrequent species thus contribute only 1% of total abundance over the 21 years of the study. Importantly, the fish at Hinkley Point fall into two distinct groups: a core of persistent (more than 10 years in the record) and usually—but not invariably—abundant species and a set of occasional (less than 10 years in the record) and typically non-abundant species (Fig. 1). This allows us to decompose the empirical species abundance distribution (Fig. 1b) into two groups of species. The distribution of the occasional species, which are not continuously present in an assemblage but may sometimes breed there, follows a log series distribution (Fig. 1d)², whereas the persistent species are log normally distributed (Fig. 1c)^{3,8}. We employ three different methods to demonstrate that the division of species into two groups is not arbitrary.

First, we use a diversity statistic to reveal the shift from the log series to the log normal model as we work through the data set. The average value of Simpson’s diversity index, D , will remain approximately constant, once $S > 10$, if species are distributed according to the log series model⁸. However, if a log normal distribution pertains, D will increase with S . We therefore expect Simpson’s index to track the transition from log series to log normal as persistent species are progressively included in the analysis. Figure 2 confirms this. The switch is to the right of the natural break in our empirical data set

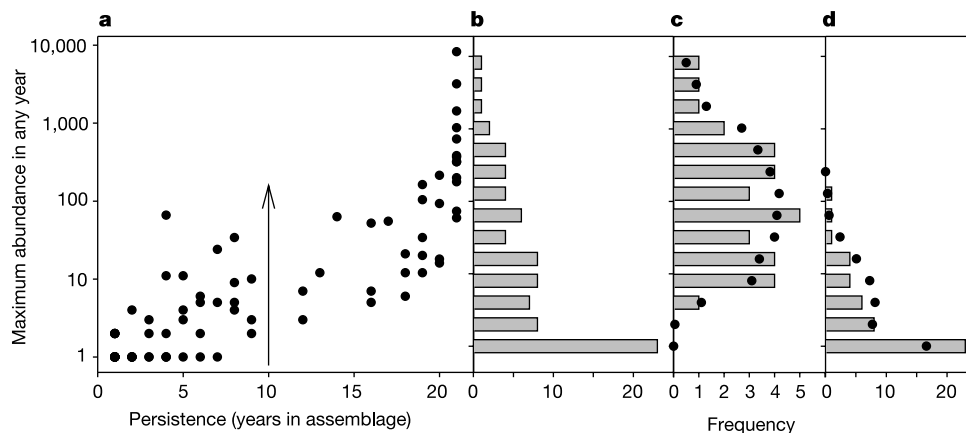


Figure 1 The pattern of abundance and persistence in the fish community of Hinkley Point, Bristol Channel. The data are for a 21-year time series of monthly samples. **a**, The number of years for which each fish was observed, plotted against the maximum abundance in any one year. A discontinuity (indicated by the vertical arrow) allows the core and occasional species to be defined as those present in >10 and <10 years.

b, The abundance distribution for all species. **c**, The abundance distribution of the core species; the frequency of each abundance class predicted with a log normal model is shown as a dot ($\chi^2_{[6]} = 0.88$, $P = 0.99$). **d**, The abundance of the occasional species; the frequency of each abundance class predicted with a log series model is shown as a dot ($\chi^2_{[4]} = 4.24$, $P = 0.39$).

because this analysis checks the signal of the underlying model as the proportion of core species increases. Thus, it is only when species that occur in 14 or more years are included that the log normal distribution drowns out the log series distribution created by the abundances of the less-persistent species.

Second, core and occasional species should be biologically distinguishable. We assigned species to habitat categories using descriptions in refs 11 and 12. Fish predominantly associated with muddy substratum, estuaries or with anadromous/catadromous life histories were allocated to the estuarine group; those that prefer rock, sand, gravel or weed substratum or are found in deeper water were placed in the non-estuarine group. Typical examples in the estuarine group are sprat, whiting, European seabass (*Dicentrarchus labrax*) and flounder (*Platichthys flesus*). Non-estuarine species include the pilchard (*Sardina pilchardus*), Ballan wrasse (*Labrus bergylta*) and 15-spined stickleback (*Spinachia spinachia*), which are typically found in deep water, rocky and weedy habitats respectively. A χ^2 test revealed a highly significant association between the biological and empirical allocation of species (estuarine core species, $n = 25$; estuarine occasional species, $n = 8$; non-estuarine core species, $n = 8$; non-estuarine occasional species, $n = 38$ (1 species not assigned); $\chi^2 = 26.9$, $P < 0.001$).

Last, the arrival of occasional species can be modelled as a stochastic event often related to unusual weather conditions such as storms or unusually settled and warm periods. However, not all arrivals are independent of each other, because particular climatic conditions will favour certain categories of fish. Overall, a Poisson process can approximate the pattern of arrival of each of these species. It was this process that Fisher invoked when he proposed his log series model. If a species follows a Poisson distribution, the variance to mean ratio will not be significantly different from 1.0 (ref. 13). Figure 3 demonstrates that this ratio successfully separates

the species into the persistent and occasional categories previously identified by using the empirical species abundance distribution. As long as a reasonable time series is available, this approach provides a tractable and objective solution to the difficult challenge of fitting a mixture of two distributions to species abundance data. Occasional species can be identified by using the variance/mean ratio, then assigned to the log series distribution and removed, and a log normal fitted to the remaining species distribution.

Although the precise location of the split point varies depending on the method used, each of these independent approaches is consistent in demonstrating that the species that comprise the empirical distribution fall into two categories. In practice, the exact position of the division is relatively unimportant from a model-fitting perspective because both the log series and the log normal are robust against moderate shifts in relative abundance^{14,15}. As the time series lengthens we expect the modes of the two distributions to move apart and the division zone to become broader and clearer. Snapshot surveys, which record species abundances at one point in time, obscure these complexities.

Southwood¹⁶ noted that terrestrial insect assemblages are continually challenged by a flow of migrants and suggested that the balance between transient and core species will determine whether a species abundance distribution is log normal or log series in character. Hubbell⁶ also concluded that the distribution of species abundance is a function of immigration rate. Our results extend these insights by showing that the occasional and persistent species leave different signatures on the species abundance distribution. However, we depart from Hubbell on the neutrality argument. As we have shown, there are clear differences in the ecological requirements of species in the two components of the distribution. Explanations of species abundance distributions cannot therefore be divorced from biology^{4,7,9}. Species are also lost over longer time scales: in our study area the eel (*Anguilla anguilla*) is currently declining and might shortly become an occasional visitor rather than a member of the core community. These longer-term changes might explain why the log normal distribution of the core species retains some negative skew.

Brown *et al.*¹⁷ contend that systems open to colonization are better able to withstand natural or human induced perturbations because immigrants can replace missing species and tolerate changed conditions. This is why islands, with their reduced flow

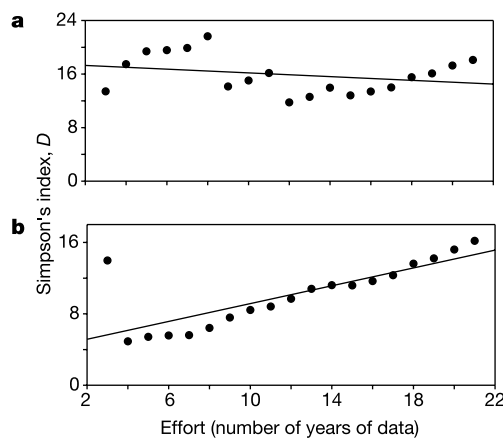


Figure 2 Switch from log series to log normal model revealed by Simpson's index. We calculated the value of Simpson's index, D (reciprocal form¹⁴), in relation to increasing effort (number of years of data) for 10 progressively larger subsets of the fauna. These were chosen to span the transition between the occasional and core species identified by the empirical species abundance distribution. **a**, D remained constant with effort when calculated for all species found in 13 or fewer years in the record ($F_{1,16} = 2.98$, slope, $b = -0.212$, $P > 0.1$). This is consistent with a log series distribution⁸. **b**, When the analysis was run with species present in 14 or fewer years, a strong positive relationship, indicative of a log normal model⁹, emerged ($F_{1,16} = 1105.9$, $P < 0.001$, $b = 0.66$), confirming the shift in species abundance distribution. Analyses that included species with longer or shorter representation in the time series gave consistent results (≤ 6 years: $F_{1,16} = 4.85$, $P = 0.04$, $b = -0.315$; ≤ 7 years: $F_{1,16} = 3.78$, $P = 0.07$, $b = -0.266$; ≤ 8 years: $F_{1,16} = 3.75$, $P = 0.07$, $b = -0.193$; ≤ 9 years: $F_{1,16} = 3.34$, $P = 0.09$, $b = -0.217$; ≤ 12 years: $F_{1,16} = 3.61$, $P = 0.07$, $b = -0.229$; ≤ 16 years: $F_{1,16} = 171.2$, $P < 0.001$, $b = 0.40$; ≤ 17 years: $F_{1,16} = 326.8$, $P < 0.001$, $b = 0.40$; ≤ 18 years: $F_{1,16} = 473.2$, $P < 0.001$, $b = 0.49$).

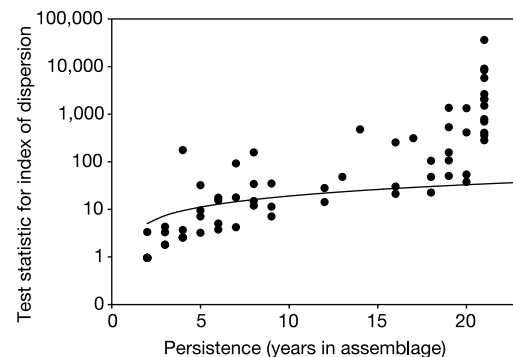


Figure 3 Using the variance/mean ratio to decompose the distribution. In accordance with ref. 13 we calculated, for each species, an index of dispersion based on the ratio of the variance to the mean. This is plotted against the number of years for which the species was present in the assemblage. The line shows the 2.5% confidence limit for the χ^2 distribution: species that fall below this line follow a Poisson distribution, and are randomly dispersed in time. Those that lie above the line are not. The 14 species that occurred in a single year are excluded from this analysis because their dispersion in time has no variance. However, these species would, by Fisher's logic, be assigned to the log series distribution. Nine species in our occasional grouping were mis-classified by this analysis. These species were found in clusters of years and their incidence increases close to our empirical split point. Only three core species are Poisson distributed.

of colonists, are especially vulnerable to environmental change. Marine and estuarine assemblages are already affected by climate change; this data set tracks the recent expansion of range and population size in fish that are close to the northern limit of their range in British waters such as sole (*Solea solea*¹⁸), seabass¹⁹ and trigger fish (*Balistes carolinensis*). Conversely, it also shows the decline in species that reach the southern limit of their range in British waters, for example sea snail (*Liparis liparis*²⁰) and dab (*Limanda limanda*²¹). If conditions alter sufficiently we predict that new core species, drawn from the pool of occasional species, will replace the existing ones. Temporal components of species abundance also have important implications for conservation planning, as recognized by the latest generation of reserve selection algorithms, which incorporate information on species permanence²². □

Methods

Fish samples were collected from the cooling-water filter screens at Hinkley Point B Nuclear Power Station, situated on the southern bank of the Bristol Channel in Somerset, England. The power station intakes are placed in front of a rocky promontory within Bridgwater Bay; to the east are the extensive Stert mud flats with an intertidal area of ~40 km². The water intakes are placed between -1 and -5 m MLWS (mean low water springs), so the fish are sampled from water of between 8 and 18 m depth. Full descriptions of the intake configuration and sampling methodology are given in refs 18 and 23. Quantitative sampling began in 1980 when 24-h surveys of the diurnal pattern of capture were undertaken in October and November. From these surveys it was concluded that samples collected during daylight were representative of the 24-h catch²⁴, and monthly quantitative sampling began in January 1981. The total volume of water sampled per month, which has not varied over the entire 21-year period, is 3.24×10^5 m³. To standardize for tidal influence, all sampling dates are chosen for tides halfway between springs and neaps, with sampling starting at high water (normally about 12:00). Fish are collected hourly from two filter screens for a 6-h period, identified to species, measured and the number of individuals recorded. Since 1987 the standard lengths (SLs) of all captured fish have been recorded to the nearest millimetre. The filter screens have a solid square mesh of 10 mm and start to retain fish >25 mm SL. A 100% retention for many species occurs at SLs > 40 mm. For fish such as sprat, whiting and pout the screens retain all fish captured with a SL greater than ~60 mm (ref. 25). The sampling method therefore catches adults and juveniles older than 6 months for all known British marine fish.

Received 25 October 2002; accepted 7 March 2003; doi:10.1038/nature01547.

1. Motomura, I. On the statistical treatment of communities [in Japanese]. *Zool. Mag., Tokyo* **44**, 379–383 (1932).
2. Fisher, R. A., Corbet, A. S. & Williams, C. B. The relation between the number of species and the number of individuals in a random sample of an animal population. *J. Anim. Ecol.* **12**, 42–58 (1943).
3. Preston, F. W. The commonness, and rarity, of species. *Ecology* **29**, 254–283 (1948).
4. Brown, J. H. Towards a general theory of biodiversity. *Evolution* **55**, 2137–2138 (2001).
5. Nee, S., Harvey, P. H. & May, R. M. Lifting the veil on abundance patterns. *Proc. R. Soc. Lond. B* **243**, 161–163 (1991).
6. Hubbell, S. P. *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton Univ. Press, Princeton, New Jersey, 2001).
7. Gaston, K. J. & Blackburn, T. M. *Macroecology* (Blackwell Science, Oxford, 2000).
8. May, R. M. In *Ecology and Evolution of Communities* (eds Cody, M. L. & Diamond, J. M.) 81–120 (Harvard Univ. Press, Cambridge, Massachusetts, 1975).
9. Sugihara, G. Minimal community structure: an explanation of species abundance patterns. *Am. Nat.* **116**, 770–787 (1980).
10. Harte, J., Kinzig, A. & Green, J. Self-similarity in the distribution and abundance of species. *Science* **284**, 334–336 (1999).
11. Lythgoe, J. & Lythgoe, G. *Fishes of the Sea* (Blandford, London, 1971).
12. Wheeler, A. *The Fishes of the British Isles* (Macmillan, London, 1969).
13. Krebs, C. J. *Ecological Methodology*, 2nd edn (Harper & Row, New York, 1999).
14. Magurran, A. E. *Ecological Diversity and its Measurement* (Princeton Univ. Press, Princeton, New Jersey, 1988).
15. Southwood, R. & Henderson, P. A. *Ecological Methods* (Blackwell Science, Oxford, 2000).
16. Southwood, T. R. E. The Croonian Lecture 1995. Natural communities: structure and dynamics. *Phil. Trans. R. Soc. Lond. B* **351**, 1113–1129 (1996).
17. Brown, J. H., Whitam, T. G., Ernest, S. K. M. & Gehring, C. A. Complex species interactions and the dynamics of ecological systems: long term experiments. *Science* **293**, 643–650 (2001).
18. Henderson, P. A. & Seaby, R. M. H. On the factors influencing juvenile flatfish abundance in the lower Severn Estuary. *Neth. J. Sea Res.* **32**, 321–330 (1994).
19. Henderson, P. A. & Corps, M. The role of temperature and cannibalism in interannual recruitment variation of bass in British waters. *J. Fish Biol.* **50**, 280–295 (1997).
20. Henderson, P. A. & Seaby, R. M. H. Population stability of the sea snail at the southern edge of its range. *J. Fish Biol.* **54**, 1161–1176 (1999).
21. Henderson, P. A. On variation in the dab, *Limanda limanda*, recruitment: a zoogeographic study. *J. Sea Res.* **40**, 131–142 (1988).
22. Rodrigues, A. S. L., Gaston, K. J. & Gregory, R. D. Using presence-absence data to establish reserve selection procedures that are robust to temporal species turnover. *Proc. R. Soc. Lond. B* **267**, 897–902 (2000).
23. Henderson, P. A. & Holmes, R. H. A. On the population dynamics of dab, sole and flounder within Bridgwater bay in the lower Severn Estuary, England. *Neth. J. Sea Res.* **27**, 337–344 (1991).

24. Henderson, P. A. & Holmes, R. H. A. Population stability over a ten-year period in the short-lived fish *Liparis liparis* (L.). *J. Fish Biol.* **37**, 605–616 (1990).
25. Turnpenny, A. W. H. *The use of multiple regression analysis in formulating operating strategies for power impingement problems* (CEGB publication RD/L/2043N81, Central Electricity Generating Board, Leatherhead, 1981).

Acknowledgements We thank R. Seaby and R. Somes for assistance with fieldwork.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.E.M. (e-mail: aem1@st-andrews.ac.uk). The full data set can be downloaded from the Pisces Conservation Ltd website (<http://www.irchouse.demon.co.uk/latestreports.html>).

Decisions about parental care in response to perceived paternity

Bryan D. Neff

Department of Biology, University of Western Ontario, London, Ontario N6A 5B7, Canada

Evolutionary ecologists are attempting to explain how parents make behavioural decisions about how much care to provide to their young^{1–4}. Theory predicts that when genetic relatedness to young is decreased by cuckoldry, for example, parents should reduce their care in favour of alternative broods that provide greater reproductive success^{5–7}. Experimental manipulation of perceived paternity has been used to test the theory^{8,9}, but such studies have generated mixed results^{10–13}. Some manipulations can fail to alter a parent's perceived paternity¹⁴, whereas others may directly affect parental behaviour when, for instance, the manipulation involves capturing the parent^{15–18}. No study has demonstrated parental care adjustment in a manner uncomplicated by experimental design or life history correlates. Here I test the theory using the fact that nest-tending parental male bluegill sunfish (*Lepomis macrochirus*) can assess their paternity using both the visual presence of parasitic cuckolder males during spawning¹⁹, and olfactory cues released by newly hatched eggs^{20,21}. By manipulating both types of cues I show that parental males dynamically adjust their parental care, favouring broods that are apparently most closely related. These results confirm the importance of genetic relatedness in parental care decision-making.

Bluegill are native to lakes and rivers of North America. Males are characterized by a discrete polymorphism in life histories termed 'parental' and 'cuckolder'^{22,23}. In Lake Opinicon (Ontario, Canada), parentals mature at age 7 years and construct nests in a colony during the breeding season²³. Nesting parentals court and spawn with multiple females over the course of a single day and then provide sole care for the developing young in their nests. Parental care involves fanning and defending eggs (which can number in the tens of thousands) until they hatch (2–3 days), and then defending developing fry from predators until the young leave the nest (5–7 days). If a parental abandons his nest before this time, the young do not survive. Parentals do not forage while they are tending their brood and so they lose about ten per cent of their body weight, although they do occasionally cannibalize some of the eggs or fry in their nest²⁴. After the fry have dispersed, parentals return to deeper waters to feed and replenish their energy reserves before re-nesting in a subsequent spawning bout²⁵.

In contrast, cuckolders do not build nests or care for young but instead mature precociously and steal fertilizations in the nests of

parentals using two age-dependent tactics. 'Sneakers' (age 2–3 yr) hide behind plants and debris near the nest edge but are visible as they dart into the nest to release sperm when a female deposits her eggs. 'Satellites' (age 4–5 yr) are about the size of mature females (age 4–8 yr) and they exhibit female colouration and behaviour, leading parentals to misidentify them as a second female in the nest^{19,22}. Thus, parentals can use the presence of sneakers during spawning as an indirect cue of reduced paternity^{19,26}. In addition, once the eggs hatch the developing fry release a direct (chemical) cue, possibly in their urine²⁷, that parentals can use to distinguish their offspring from those of both sneakers and satellites^{19–21}.

I manipulated the perceived paternity of nesting parental males to see how it affected their guarding and cannibalistic behaviours using two experiments (see Methods). In both, the parental's actual paternity was revealed by olfactory cues released by the fry^{20,21}. These cues were not present during the egg phase of care²⁰. The strength of the two experiments lies in their opposite predictions with respect to the change in parental care from the egg phase to the fry phase. In the first, treatment parentals (fish whose nests were manipulated) had a lower perceived paternity during the egg phase of care, but a higher perceived paternity during the fry phase, whereas in the second, treatment parentals had a higher perceived paternity during the egg phase of care than during the fry phase.

In the first experiment, four sneakers (young cuckolders) were placed inside small transparent plastic containers near the nests of 34 randomly selected parentals within a colony for the duration of spawning. This manipulation should have reduced parentals' perceived paternity because proximity of sneakers to the nest during spawning is a reliable threat to paternity²⁶. Six of these males abandoned their nests shortly after spawning, while an additional two abandoned nests shortly after the eggs hatched. Empty containers were similarly placed around 20 other 'control' nests. In this case, two males abandoned their nests shortly after spawning and three abandoned nests after the eggs hatched. The day after spawning, male behaviour was quantified by evaluating each parental's willingness to defend his brood from a potential egg predator (a pumpkinseed sunfish in a clear plastic container: see Methods). The results showed that egg defence was significantly lower by the treatment males than the control males (excluding males that abandoned nests: $t_{39} = 2.7$, $P = 0.011$; Fig. 1a).

Sneakers in the plastic containers were prevented from siring any eggs, so after the eggs of experiment 1 males hatched, the parental males should have been able to use the olfactory cues released by the fry to reassess their actual paternity (and discover that they had been cuckolded less than anticipated). As predicted, after the eggs

hatched there was no difference in the parental care of the treatment and control groups of males ($t_{39} = 0.64$, $P = 0.52$; Fig. 1a). This test assumes that 'natural' cuckoldry rates were not significantly higher in control nests. Examining the change in parental care between the egg and fry phases, treatment males increased their level of care significantly more than control males ($t_{39} = 2.79$, $P = 0.008$; Fig. 2).

In the second experiment, the day after spawning, approximately one-third of the eggs were removed from nests of 20 parentals and replaced with unrelated eggs from one of the other males' nests. Four of these males abandoned their nests shortly after the eggs hatched. Sham-swaps were performed on eggs of 15 control parentals, but one of these males abandoned the nest shortly after the swap, and another abandoned his shortly after the eggs hatched. The parental behaviour of each male was quantified three times: before the egg swap (to provide a baseline), the day after the swap, and the day after the eggs hatched. There was no difference in the willingness of treatment or control parental males to defend their nests prior to ($t_{27} = 0.97$, $P = 0.34$) or the day after the manipulation ($t_{27} = 1.22$, $P = 0.23$) (Fig. 1b). Both groups did increase their level of care between these two periods (paired t -test for control: $t_{12} = 5.60$, $P < 0.001$; for treatment: $t_{15} = 5.22$, $P < 0.001$), but the change was similar ($t_{27} = 0.37$, $P = 0.71$; Fig. 2). Thus, if the manipulation affected the parental males, it did so equivalently.

However, after the eggs hatched, there was a significant difference between the two groups—treatment parental males provided less care than controls ($t_{27} = 2.08$, $P = 0.047$; Fig. 1b). This occurred because treatment males decreased their level of care significantly more than control males ($t_{27} = 4.58$, $P < 0.001$; Fig. 2). This test is powerful because it controls for natural variation in the level of care among parentals within the treatment and control groups. Within-group variation may be attributed to, for example, differences in condition¹⁹. Furthermore, relative to the baseline measure, control males increased their level of care (paired t -test: $t_{12} = 2.65$, $P = 0.021$), but treatment males decreased their level of care (paired t -test: $t_{15} = -3.30$, $P = 0.005$), and these adjustments in parental behaviour were significantly different between

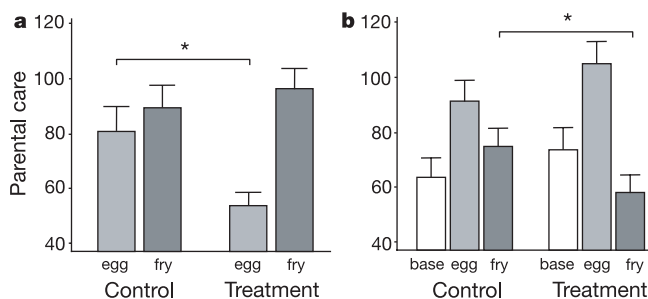


Figure 1 Parental care by control and treatment parental male bluegill. **a**, Results from the sneaker manipulation (experiment 1) showed that the parental care of treatment males was significantly lower than control males during the egg phase but not during the fry phase of care. **b**, Results from the egg manipulation (experiment 2) showed that the parental care of treatment males was similar before the manipulation (baseline) and the day following the manipulation (egg phase), but was significantly lower than control males during the fry phase of care. Error bars represent 1 s.e. and horizontal lines with asterisks denote significant differences.

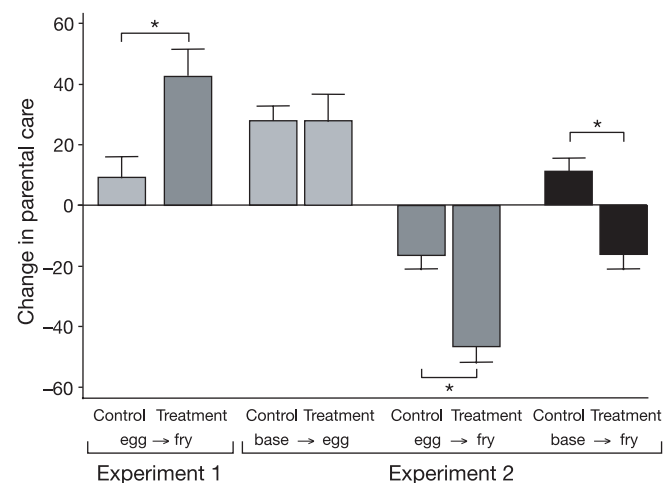


Figure 2 Changes in parental care by control and treatment parental male bluegill. In experiment 1, treatment males increased parental care more than did control males between the egg and fry phases of care (egg → fry). In experiment 2, both treatment and control males increased parental care similarly between the baseline and the day after the manipulation (base → egg). However, after the eggs hatched (egg → fry) treatment males decreased parental care more so than did control males and relative to the baseline (base → fry) these latter levels reflect a decrease for treatment males and an increase for control males. Error bars represent 1 s.e. and horizontal lines with asterisks denote significant differences.

the two groups ($t_{27} = 4.41$, $P < 0.001$; Fig. 2).

Further support comes from three additional analyses. First, in experiment 1, the manipulation should have led to greater abandonment shortly after spawning in the treatment than control group, whereas in experiment 2 abandonment should have been greater in the treatment than control group shortly after the eggs hatched. Combining data from both experiments, 10 treatment males (6 in experiment 1 and 4 in experiment 2) and 3 control males (2 in experiment 1 and 1 in experiment 2) abandoned nests during these predicted times. The probability that 10 or more males would abandon from the treatment group yet 3 or fewer from the control group is unlikely to occur by chance (cumulative binomial probability: $P = 0.065$).

Second, as expected in experiment 1, there was no difference in the pecking behaviour (indicative of partial brood cannibalism) of treatment and control males during the fry phase of care (treatment: 5.7 ± 0.6 (s.e.); control: 7.5 ± 1.5 ; $t_{39} = 1.34$, $P = 0.19$). However, in experiment 2 treatment males did peck at their fry more often, although the difference was not significant (treatment: 15.3 ± 1.8 ; control: 10.8 ± 1.6 ; $t_{27} = 1.81$, $P = 0.081$).

Third, I compared the changes in parental care between the baseline and fry phase in experiment 2 to those made by parentals in response to natural variation in perceived paternity observed in a previous correlative study¹⁹. On the basis of those data¹⁹, the relative level of parental care C_{rel} between the egg and fry phases ($C_{rel} = C_{fry}/C_{egg}$) was related to the change in an individual's perceived paternity ($\Delta P_{per} = P_{fry} - P_{egg}$) (see Fig. 1 in ref. 19) according to the following equation: $C_{rel} = 10^{(0.078 + 0.873 \times \Delta P_{per})}$. Thus, my control males should have had a C_{rel} of 1.20 ($\Delta P_{per} = 0$). They in fact had a very similar C_{rel} of 1.18 ($=75.6/63.9$). For the treatment males, ΔP_{per} should have averaged -26% ($= -\frac{1}{3} \times 79\%$, where the proportion of eggs that were swapped was one-third and 79% is the average paternity of parentals in their nests²⁸) and therefore they should have had a C_{rel} of 0.71. Although the observed C_{rel} of 0.78 ($=57.8/74.3$) is higher than this expected value, the observed value suggests that I swapped an average of 27% of the eggs, which is only marginally fewer than the estimated 33.3%.

My results cannot be due to natural variations in life history (for example, male quality, condition, age) or ecology (for example, nest location, vegetation, refugia) because parental males were randomly assigned to treatment and control groups, and only perceived paternity was experimentally manipulated. Indeed, there were no differences in the size, condition, number of eggs, or nest location of treatment and control males ($P > 0.16$ for each). It is possible that in the first experiment, treatment males expended more energy during spawning by attempting to chase away the sneakers in the containers and had less energy to invest in brood defence, or perhaps thought that these sneakers were brood predators (although sneakers are much smaller than typical brood predators²³). However, these explanations would not predict the observed change in parental care made by the treatment males between the egg and fry phases relative to the control males. Furthermore, the second experiment employed analogous, simultaneous manipulations in both treatment and control nests. Collectively, these results provide compelling support for a fundamental prediction of parental care theory: parents invest care according to the evolutionary value (genetic relatedness) of their young. □

Methods

Study site

Bluegill spawning activity was monitored at Lake Opinicon (44° 34' N, 76° 19' W) during the May–July breeding season in 2000. Lake Opinicon is an 890-hectare, mesotrophic, warm-water lake, which sustains a large natural population of bluegill whose reproductive biology has been studied since the mid-1970s²³. For this study, reproductive activity was examined by swimmers equipped with snorkelling gear in an area bounding 2 km of littoral zone along the northern shore of the lake. Surveys were made daily beginning before the onset of breeding in May and ending after breeding ceased in July. During this time, two breeding colonies were selected for the manipulations and each nest was tagged

with a small numbered tile. Each colony was mapped and the position of each nest was recorded. Control and treatment nests were then randomly assigned. At the conclusion of the experiment, each parental was briefly captured to measure total body length (to the nearest mm) and weight (to the nearest 0.1 g). Fulton's condition was then calculated from weight divided by cubic length.

Experiment 1

On the morning of spawning (identified by the presence of a large school of females near the colony at about 08:00 Eastern Standard Time, EST), two sneakers were placed inside each plastic container (20 × 15 × 10 cm), and two containers were then placed on either side of each treatment nest within the colony for the duration of spawning (09:00–17:00). Thus, there were a total of four sneakers in proximity to treatment nests at all times. Sneakers ranged in size from 71–101 mm and averaged 80 mm. Control nests had two empty containers placed around the nest edge. The empty containers appeared not to interfere with spawning, while the containers with sneakers occasionally solicited an aggressive approach by the nest-tending parental. Sneakers reacted to the approach by manoeuvring to the bottom of the container where they were less conspicuous. When spawning concluded all containers were removed from the colony. The following day (09:00–11:00), the parental care of each male was quantified using established methods (see below) and an egg number score was recorded for each nest as a rank between 1 (few eggs) and 5 (many eggs)²⁵. Parental care was again recorded the day after the eggs hatched, and the following day the pecking behaviour was quantified (see below).

Experiment 2

The day after spawning (09:00–11:00), egg number scores were recorded and a baseline level of parental care was calculated for each male. Approximately one-third of the eggs from the nests of experimental parental males were then swapped. Swaps were performed between nests that had equivalent egg scores, and were always between non-neighbouring parentals to ensure the foreign eggs introduced were unrelated to the focal parental²⁸. Sham-swaps were performed in each of the control nests by removing a third of the eggs, swimming away from the nest and then returning to the nest and replacing the same eggs. Parental care was retested the day following the manipulations, and a third time the day after the eggs hatched. Two days after egg hatching, the pecking behaviour was quantified.

Parental care

Parental care was quantified using brood defence by presenting a live brood predator (pumpkinseed sunfish, *Lepomis gibbosus*) in a clear bag at the edge of each parental male's nest¹⁹. A trial consisted of presenting the predator for 30 s, removing it for 30 s, and then presenting the predator for another 30 s. Control and experimental males were tested alternately. The experimental status of each nest was unknown to the observer. An index of the parental's willingness to defend his brood was later calculated from the equation: Brood defence = $1 \times L + 2 \times F + 3 \times B$; where L , F and B are the total number of lateral displays, opercular flares and bites performed by the parental male during the trial. The coefficients were selected to reflect the relative intensity of the parental's reaction and the potential for personal injury^{19,24}. Brood defence should be a good estimate of actual parental investment.

Pecking behaviour

Parentals will occasionally 'peck' at their nest and presumably consume some of their brood (partial cannibalism)^{20,23}. A peck is an obvious behaviour whereby a male angles downward towards his nest, followed by forward movement towards the base of the nest, where the eggs or fry reside. The total number of pecks on a brood was counted for each parental based on two 30-min observation periods: one in the morning and one in the afternoon. Typically, four nests were simultaneously observed. It was not possible to record pecking behaviours during the egg phase of care owing to time constraints with the defence measurements. However, during the fry phase there was consistency between the number of pecks and each male's fry defence score ($r = -0.35$, $P = 0.003$, $n = 70$).

Received 11 December 2002; accepted 19 February 2003; doi:10.1038/nature01528.

- Westneat, D. F. & Sargent, R. C. Sex and parenting: the effects of sexual conflict and parentage on parental strategies. *Trends Ecol. Evol.* **11**, 87–91 (1996).
- Sargent, R. C. & Gross, M. R. in *The Behaviour of Teleost Fishes* (ed. Pitcher, T. J.) 275–293 (Chapman & Hall, New York, 1993).
- Burke, T., Davies, N. B., Bruford, M. W. & Hatchwell, B. J. Parental care and mating behaviour of polyandrous dunlocks *Prunella modularis* related to paternity by DNA fingerprinting. *Nature* **338**, 249–251 (1989).
- Dixon, A., Ross, D., Omalley, S. L. C. & Burke, T. Paternal investment inversely related to degree of extra-pair paternity in the reed bunting. *Nature* **371**, 698–700 (1994).
- Trivers, R. L. in *Sexual Selection and the Descent of Man* (ed. Campbell, B.) 136–179 (Aldine, Chicago, 1972).
- Westneat, D. F. & Sherman, P. W. Parentage and the evolution of parental behavior. *Behav. Ecol.* **4**, 66–77 (1993).
- Kokko, H. Cuckoldry and the stability of biparental care. *Ecol. Lett.* **2**, 247–255 (1999).
- Lessells, C. M. in *Behavioural Ecology: an Evolutionary Approach* (eds Krebs, J. R. & Davies, N. B.) 32–68 (Blackwell, Oxford, 1991).
- Kempnaers, B. & Sheldon, B. C. Studying paternity and parental care: pitfalls and problems. *Anim. Behav.* **53**, 423–427 (1997).
- Kempnaers, B., Lanctot, R. B. & Robertson, R. J. Certainty of paternity and paternal investment in eastern bluebirds and tree swallows. *Anim. Behav.* **55**, 845–860 (1998).
- MacDougall-Shackleton, E. A. & Robertson, R. J. Confidence of paternity and paternal care by eastern bluebirds. *Behav. Ecol.* **9**, 201–205 (1998).
- Sheldon, B. C. & Ellegren, H. Paternal effort related to experimentally manipulated paternity of male collared flycatchers. *Proc. R. Soc. Lond. B* **265**, 1737–1742 (1998).

13. Hunt, J. & Simmons, L. W. Confidence of paternity and paternal care: covariation revealed through the experimental manipulation of the mating system in the beetle *Onthophagus taurus*. *J. Evol. Biol.* **15**, 784–795 (2002).
14. Neff, B. D. & Sherman, P. W. Decision making and recognition mechanisms. *Proc. R. Soc. Lond. B* **269**, 1435–1441 (2002).
15. Jamieson, I. G. & Quinn, J. S. Problems with removal experiments designed to test the relationship between paternity and parental effort in a socially polyandrous bird. *Auk* **114**, 291–295 (1997).
16. Wright, J. in *Sperm Competition and Sexual Selection* (eds Birkhead, T. R. & Møller, A. P.) 117–145 (Academic, London, 1998).
17. Kokko, H. & McRae, S. B. Take care when studying parenting behaviour. *Trends Ecol. Evol.* **15**, 440–441 (2000).
18. Sheldon, B. C. Relating paternity to paternal care. *Phil. Trans. R. Soc. Lond. B* **357**, 341–350 (2002).
19. Neff, B. D. & Gross, M. R. Dynamic adjustment of parental care in response to perceived paternity. *Proc. R. Soc. Lond. B* **268**, 1559–1565 (2001).
20. Neff, B. D. & Sherman, P. W. Nestling recognition via direct cues by parental male bluegill sunfish (*Lepomis macrochirus*). *Anim. Cogn.* (in the press).
21. Neff, B. D. & Sherman, P. W. Kin recognition via self-referencing by a fish with paternal care and cuckoldry. *Proc. R. Soc. Lond. B* (submitted).
22. Dominey, W. J. Female mimicry in male bluegill sunfish—a genetic polymorphism? *Nature* **284**, 546–548 (1980).
23. Gross, M. R. Sneakers, satellites and parental males: polymorphic mating strategies in North American sunfishes. *Z. Tierpsychol.* **60**, 1–26 (1982).
24. Coleman, R. M. & Fischer, R. U. Brood size, male fanning effort and the energetics of a non-shareable parental investment in bluegill sunfish, *Lepomis macrochirus* (Teleostei: Centrarchidae). *Ethology* **87**, 177–188 (1991).
25. Cargnelli, L. M. & Gross, M. R. The temporal dimension in fish recruitment: birth date, body size, and size-dependent survival in a sunfish (bluegill: *Lepomis macrochirus*). *Can. J. Fish. Aquat. Sci.* **53**, 360–367 (1996).
26. Fu, P., Neff, B. D. & Gross, M. R. Tactic-specific success in sperm competition. *Proc. R. Soc. Lond. B* **268**, 1105–1112 (2001).
27. Brown, G. E. & Brown, J. A. Kin discrimination in salmonids. *Rev. Fish Biol. Fish.* **6**, 201–219 (1996).
28. Neff, B. D. Genetic paternity analysis and breeding success in bluegill sunfish (*Lepomis macrochirus*). *J. Hered.* **92**, 111–119 (2001).

Acknowledgements I thank L. M. Cargnelli, M. R. Gross, E. A. MacDougall-Shackleton, J. D. Olden, T. E. Pitcher and P. W. Sherman, and H. Kokko, J. D. Reynolds and B. C. Sheldon for comments on this work. I am indebted to P. Fu for providing field assistance. This work was conducted at the Queen's University Biological Station and conformed to protocols outlined by the Canadian Council on Animal Care. The research was supported by the Sigma Xi and the Natural Sciences and Engineering Research Council of Canada.

Competing interests statement The author declares that he has no competing financial interests.

Correspondence and requests for materials should be addressed to B.D.N. (e-mail: bniff@uwyo.ca).

Adaptation of photoperiodic control pathways produces short-day flowering in rice

Ryosuke Hayama^{*,†}, Shuji Yokoi^{*}, Shojiro Tamaki^{*}, Masahiro Yano[‡] & Ko Shimamoto^{*}

^{*} Laboratory of Plant Molecular Genetics, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma 630-0101, Japan

[‡] National Institute of Agrobiological Sciences, Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan

The photoperiodic control of flowering is one of the important developmental processes of plants because it is directly related to successful reproduction¹. Although the molecular genetic analysis of *Arabidopsis thaliana*, a long-day (LD) plant, has provided models to explain the control of flowering time in this species^{2–4}, very little is known about its molecular mechanisms for short-day (SD) plants. Here we show how the photoperiodic control of flowering is regulated in rice, a SD plant. Overexpression of *OsGI*, an orthologue of the *Arabidopsis GIGANTEA (GI)* gene^{6,7} in transgenic rice, caused late flowering under both SD and LD

conditions. Expression of the rice orthologue⁸ of the *Arabidopsis CONSTANS (CO)* gene⁹ was increased in the transgenic rice, whereas expression of the rice orthologue¹⁰ of *FLOWERING LOCUS T (FT)*^{11,12} was suppressed. Our results indicate that three key regulatory genes for the photoperiodic control of flowering are conserved between *Arabidopsis*, a LD plant, and rice, a SD plant, but regulation of the *FT* gene by *CO* was reversed, resulting in the suppression of flowering in rice under LD conditions.

We previously isolated *OsGI*, a rice orthologue of the *Arabidopsis GI*, by a differential display method⁵. We found that *OsGI* expression was suppressed in the photoperiod-insensitive *se5* mutant¹³, indicating that it might have a role in inhibiting the flowering of rice under LD⁵. We also found that expression of the *OsGI* messenger RNA was circadian-controlled and that its temporal expression pattern was very similar to that of *GI* under both SD and LD conditions⁵. To further understand the role of *OsGI* in the photoperiodic control of flowering in rice, we fused this gene with a gene promoter that gives constitutively high expression in rice and introduced it into rice by *Agrobacterium*-mediated transformation. The analysis of flowering time for three independent transgenic lines clearly indicated that the flowering times of these lines were later than those of the wild type both under SD and LD conditions (Fig. 1). Particularly under SD conditions, line 18 showed a markedly late phenotype, about 90 days later than that of the wild type. Even under LD conditions, which are suppressive conditions for the flowering of rice, a significant delay of flowering was observed. These results suggested that *OsGI* acts as a suppressor of flowering in rice, which is a reversal of the role of *GI* in the photoperiodic control of flowering in *Arabidopsis*^{6,14}.

The rice orthologues of the two key regulators of flowering time in *Arabidopsis*, *CO* and *FT*, were recently shown to be important in the photoperiodic control of flowering in rice^{8,10}. The rice *Hd1(Se1)* gene, which is an orthologue of the *Arabidopsis CO* gene, was required for the suppression of flowering under LD conditions and for the promotion of flowering under SD conditions⁸. Furthermore, the rice *Hd3a* gene, an orthologue of the *Arabidopsis FT* gene, was shown to be an activator of flowering in rice¹⁰. It was previously demonstrated that *CO* controls the flowering time by integrating signals from the circadian clock and light to regulate *FT* expression in *Arabidopsis*^{15–17}. Because *GI* was shown to function as an activator of *CO*¹³, we measured the mRNA levels of *Hd1(Se1)* (rice *CO* orthologue) and *Hd3a* (rice *FT* orthologue) in transgenic rice overexpressing *OsGI* (Fig. 2; Supplementary Fig. 1). The results of

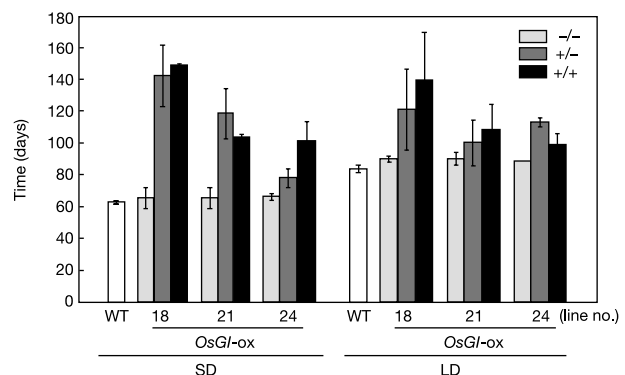


Figure 1 Flowering times of transgenic rice plants overexpressing *OsGI*. Segregating T₁ generation and wild-type (WT) plants were grown under LD and SD conditions, and the number of days to heading (flowering) was measured. In each transgenic line, 10–15 plants were used. Southern blot analysis was performed to determine the genotype of each plant and classify them into groups (+/+, homozygous for transgene; +/-, heterozygous for transgene; -/-, no transgene).

[†] Present address: Max Planck Institute for Plant Breeding Research, Carl-von-Linné-Weg 10, D-50829 Köln, Germany.

the analysis showed that *Hd1(Se1)* mRNA levels were increased under both SD and LD conditions and that the *OsGI* mRNA levels were positively correlated with those of *Hd1(Se1)* in transgenic rice (Fig. 2a and b). In contrast, the *Hd3a* mRNA levels were negatively correlated with those of *OsGI* (Fig. 2c and d), as were those of *Hd1(Se1)* (Fig. 2e and f). These results suggested that *Hd1(Se1)* acts as a suppressor of *FT* gene expression as well as of flowering in transgenic rice.

Because the mRNA levels for the rice *OsGI*, *Hd1(Se1)* and *Hd3a* genes exhibit diurnal oscillation^{5,8}, we examined the daily rhythms of their mRNA levels (Fig. 3a–h) by using leaf RNA isolated from

wild-type and transgenic plants overexpressing *OsGI* at 47 days after germination. The expression levels of *OsGI* mRNA in transgenic line 18 were relatively uniform at most times of day under both SD and LD conditions, whereas in wild-type rice plants they showed peaks just before the dark periods under both SD and LD conditions (Fig. 3a, c and d). The levels of *Hd1(Se1)* mRNAs showed characteristic rhythms under SD and LD conditions in both untransformed wild-type plants and transgenic plants overexpressing *OsGI*. In the wild-type plants under LD conditions, the *Hd1(Se1)* mRNA level was relatively high at dawn; in contrast, the expression level under LD conditions was low around midday and became high

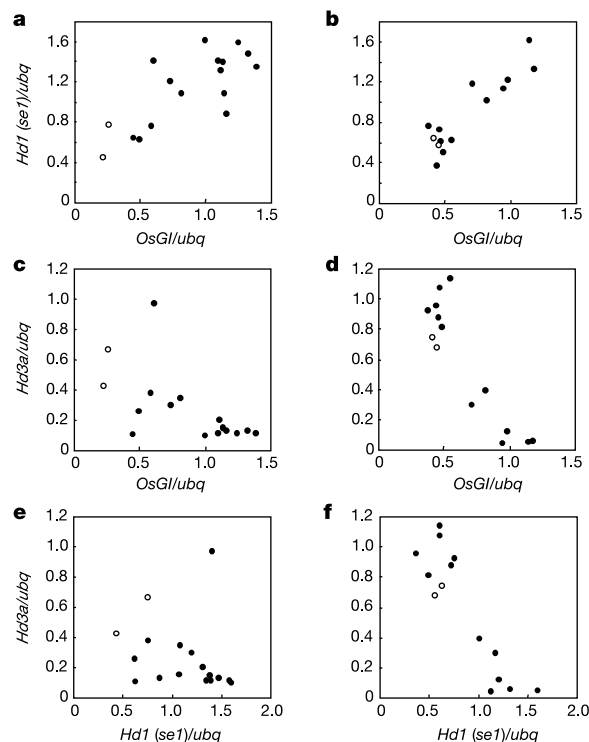
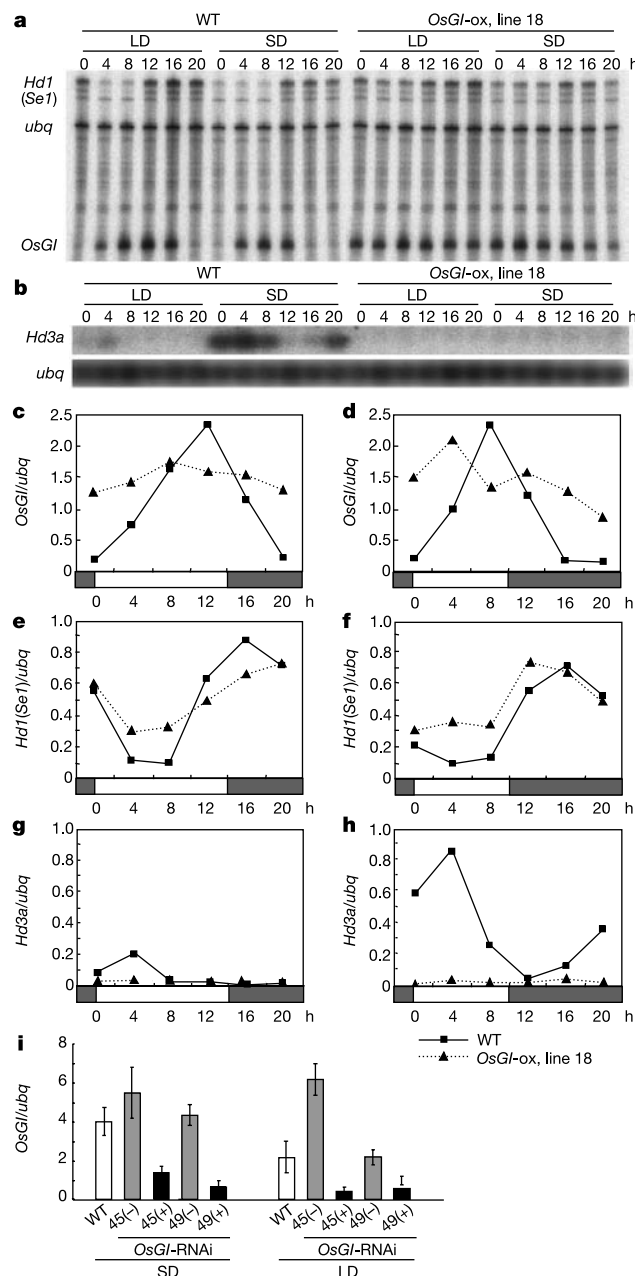


Figure 2 Correlations between *OsGI*, *Hd1(Se1)* and *Hd3a* mRNA levels in *OsGI*-ox plants. Segregating T₁ generation and wild-type plants were grown under LD and SD conditions for 52 days and leaves were collected 3 h after dawn. RT-PCR was performed and the mRNA levels of each gene were quantified from the gel images shown in Supplementary Fig. 1. Results under LD conditions are shown in **a**, **c** and **e** and results under SD conditions are shown in **b**, **d** and **f**. Control wild-type plants are indicated by open circles and transgenic plants by filled circles. **a**, **b**, Correlation between *OsGI* and *Hd1(Se1)* mRNA levels. **c**, **d**, Correlation between *OsGI* and *Hd3a* mRNA levels. **e**, **f**, Correlation between *Hd1(Se1)* and *Hd3a* mRNA levels.

Figure 3 Temporal expression patterns of *OsGI*, *Hd1(Se1)* and *Hd3a* mRNAs in an *OsGI*-ox plant (line 18) over a 24-h period and analysis of *OsGI*-RNAi plants. **a**, Detection of *Hd1(Se1)*, *OsGI* and *ubq* mRNAs by RNase protection assay. **b**, Detection of *Hd3a* and *ubq* mRNAs by RT-PCR analyses. Line 18 and the wild type (WT) were grown under LD and SD conditions for 47 days, and leaves were collected every 4 h over a 24-h period. **c**, **d**, Quantification of *OsGI* mRNA from the experiment shown in **a**, under LD (**c**) and SD (**d**) conditions. **e**, **f**, Quantification of *Hd1(Se1)* mRNA from **a**, under LD (**e**) and SD (**f**) conditions. **g**, **h**, Quantification of *Hd3a* mRNA from **b**, under LD (**g**) and SD (**h**) conditions. **i**, **j**, Quantification of *OsGI* mRNA (**i**) and flowering times (**j**; means $\pm$ s.e.m.) of the wild-type (N8) and *OsGI*-RNAi plants. Ten wild-type and 10 transgenic plants, homozygous for the transgene, and plants without transgenes at T₂ generation were grown under LD and SD conditions. Leaves were harvested 30 days after germination for RNA isolation. Four of the 10 plants were subjected to real-time PCR analysis of *OsGI* expression. Flowering times were measured for all the plants. Plus and minus signs indicate segregating T₂ transgenic plants homozygous for the transgene and those lacking the transgene, respectively.



Genotype	Days to heading	
	SD	LD
Wild type (Norin 8)	65.9 $\pm$ 0.11	115.3 $\pm$ 0.06
<i>OsGI</i> -RNAi line 45(-)	66.3 $\pm$ 0.18	117.1 $\pm$ 0.35
<i>OsGI</i> -RNAi line 45(+)	74.3 $\pm$ 1.09	108.7 $\pm$ 0.38
<i>OsGI</i> -RNAi line 49(-)	65.1 $\pm$ 0.48	115.1 $\pm$ 0.10
<i>OsGI</i> -RNAi line 49(+)	85.6 $\pm$ 0.21	107.1 $\pm$ 0.10

n = 10

during the night. Under SD conditions, the *Hd1(Se1)* mRNA level at dawn was much lower than that under LD conditions. Under LD conditions, *Hd1(Se1)* mRNA levels in transgenic plants were higher than those of the wild type during the light period, whereas under SD conditions the higher level of expression extended well into the night hours. A similar temporal pattern of *Hd1(Se1)* mRNA expression was observed in transgenic line 21 (Supplementary Fig. 2).

In contrast to the expression patterns of *GI/OsGI* and *CO/Hd1(Se1)*, *Hd3a* mRNA showed striking differences between LD and SD conditions in wild-type rice plants (Fig. 3b, g and h). It was strongly suppressed under LD conditions but showed a diurnal rhythm under SD conditions; its level was high during the day and became low during the night. No expression of *Hd3a* was detected in transgenic rice, which was consistent with their late-flowering phenotype under both LD and SD conditions (Fig. 1). Comparisons of the expression patterns of *Hd1(Se1)* between wild-type and transgenic rice under SD conditions (Fig. 3f) and between LD and SD conditions in the wild type (compare Fig. 3e with Fig. 3f) indicated that the higher expression of *Hd1(Se1)* during the light period might cause suppression of *Hd3a* mRNA expression. We

performed the same experiment by using RNA isolated from younger plants of line 18 at 30 days after germination and obtained essentially similar results (Supplementary Fig. 3). It has recently been shown that, under LD conditions, *Hd3a* mRNA expression was increased in a mutant deficient for *Hd1(Se1)*¹⁸. Our results that *Hd1(Se1)* suppressed *Hd3a* expression in *OsGI*-ox plants (see Methods) under LD conditions are consistent with this observation.

To analyse the effects of decreased *OsGI* expression on the flowering time in rice, we generated transgenic rice in which *OsGI* expression was suppressed by RNA-mediated interference (RNAi; Fig. 3i and j). The results indicated that the decreased *OsGI* expression led to early flowering under LD conditions and late flowering under SD conditions relative to the wild type, and that this early flowering phenotype under LD conditions was the opposite of that of the *Arabidopsis gi* mutants^{6,14}, confirming the opposite role of the *GI* gene in flowering under LD conditions in rice and *Arabidopsis*. The observation that the flowering phenotypes of *OsGI*-RNAi plants under both LD and SD conditions were similar to those of the *se1* mutants⁸ indicated that *OsGI* and *Hd1(Se1)* act in a similar direction in the photoperiodic control of flowering in rice. In *OsGI*-RNAi plants, *Hd1(Se1)* expression was well correlated with *OsGI* expression at 4 h after the lights were turned on, but the levels of *Hd3a* mRNA were variable and did not give a clear correlation with those of *Hd1(Se1)* mRNA (data not shown). This could be due to a smaller decrease seen in *Hd1(Se1)* mRNA in the plants carrying the *OsGI*-RNAi gene than in the segregating untransformed control plants.

To examine further the effects of *Hd1(Se1)* on *Hd3a* gene expression, we performed co-transfection of *Ubq::Hd1(Se1)* cDNA with *Hd3a::gus* in rice cell-culture protoplasts by electroporation and examined GUS activities (Fig. 4a and b). The results showed that co-transfection of *Ubq::Hd1(Se1)* cDNA with *Hd3a::gus* decreased GUS activities to 33% and 41% of those of the control in experiments 1 and 2, respectively, whereas *35S::gus* expression was not affected by co-transfection of *Ubq::Hd1(Se1)* cDNA. The expression of *Hd3a::gus* was not affected by co-transfection of a control plasmid carrying *Ubq::bar*. These results indicate that *Hd1(Se1)* can suppress the promoter activity of the *Hd3a* gene in the protoplasts, supporting a similar observation made with transgenic plants overexpressing *OsGI*. They also indicate that the cellular activities of these protoplasts might mimic those of rice leaf cells under LD conditions. This transient assay system might be a useful tool for the analysis of the molecular mechanisms of *Hd3a* suppression by *Hd1(Se1)* in rice under various environmental conditions.

Taken together, the results of the analyses of *OsGI*-ox and *OsGI*-RNAi plants and co-transfection experiments using rice protoplasts led us to propose a model to explain a gene network for the suppression of flowering under LD conditions in a SD plant, rice (Fig. 4c). In *Arabidopsis*, *GI* acts as an activator of *CO*, which in turn activates a floral activator, *FT*, under LD conditions. In rice, *OsGI* activates *Hd1(Se1)* in a similar manner to that in *Arabidopsis*. However, under LD conditions, *Hd1(Se1)* suppresses *Hd3a* expression, causing the suppression of flowering. It is nevertheless possible that there are additional regulatory mechanisms for the photoperiodic control of flowering in rice.

Our results indicate that an important gene network for the photoperiodic control of flowering is conserved between *Arabidopsis* and rice but that the regulation of the downstream gene by an upstream regulatory gene is reversed in the two species. They also identified an example in plants in which an important developmental process can be diversified by using the same set of regulatory genes but by regulating them differently. Recent findings that the introduction of the *Arabidopsis CO* gene caused the suppression of SD-induced tuberization in potato showed a reversed action of *CO* on a SD-promoted process in another plant¹⁹. Furthermore, a *CO* orthologue (*PnCO*) was isolated from another SD plant, *Pharbitis*

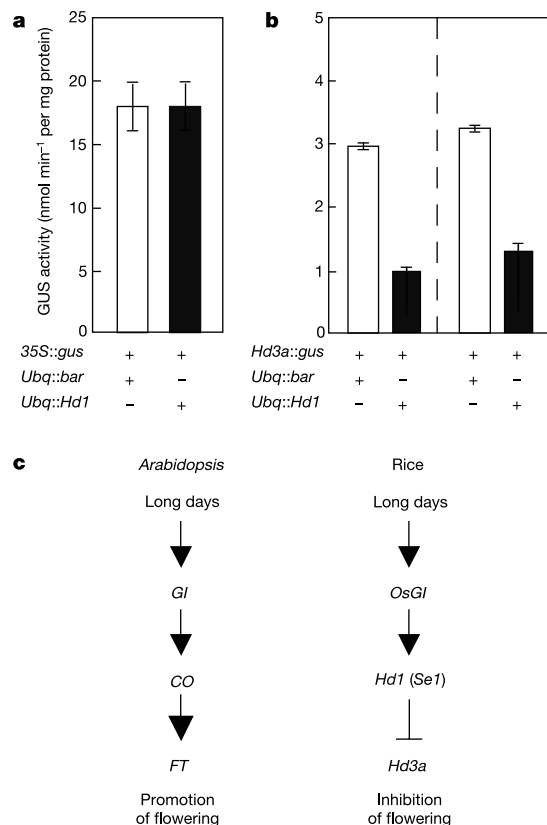


Figure 4 Transfection assays in rice protoplasts, and model of the genetic pathway controlling the photoperiodic response of flowering in rice and *Arabidopsis*. **a**, **b**, GUS activities of protoplasts transfected with the plasmid containing *35S::gus* (**a**) or *Hd3a::gus* (**b**) were measured. For each experiment, protoplasts were co-transfected with the plasmid containing *Ubq::bar* (open bar) or *Ubq::Hd1 cDNA* (filled bar). For experiment 1 in **b** (left panel), two replicates were used; for experiment 2 in **b** (right panel), six replicates were used. **c**, The molecular mechanisms for the regulation of *GI/OsGI* expression, and the regulation of *CO/Hd1(Se1)* expression by *GI/OsGI*, might be conserved between *Arabidopsis* and rice. Both the *GI-CO* and *OsGI-Hd1(Se1)* pathways are activated under LD conditions; the former promotes the flowering of *Arabidopsis*, whereas the latter inhibits the flowering of rice. The divergence in the photoperiodic pathways of these two species arises from the molecular mechanisms controlling the expression of *FT/Hd3a* by *CO/Hd1(Se1)*.

nil, and it was shown that 35S::PnCO promoted flowering in the *co* mutants of *Arabidopsis* under both LD and SD conditions²⁰. These findings suggest the existence of common mechanisms for the photoperiodic control of various processes in diverse plant species. □

Methods

Plant materials and growth conditions

Japanica rice cultivars, Norin 8, *OsGI*-ox transgenic plants and *OsGI*-RNAi transgenic plants were grown in climate chambers with 24-h temperature cycles (12 h, 30 °C during subjective day; 12 h, 25 °C during subjective night). The fluence rates of light were ~300 μmol m⁻² s⁻¹ (400–750 nm) under LD and SD conditions.

Generation of *OsGI*-ox and *OsGI*-RNAi transgenic plants

To generate *OsGI*-ox transgenic rice, *OsGI* cDNA (containing a 240-base-pair (bp) and a 274-bp fragment as 5' and 3' untranslated regions, respectively) was introduced into a Ti-based vector in which *OsGI* is driven by the maize *Ubi1* promoter²¹. To produce the hairpin RNAi construct, two 3.2-kilobase (kb) *OsGI* cDNAs were fused in reverse orientation, and a 28-bp unrelated fragment was inserted as a linker. This construct was fused with the maize *Ubi1* promoter. *Agrobacterium*-mediated transformation of rice was performed as described²². Transformed calluses were selected by hygromycin resistance, and plants were regenerated from transformed calluses.

Polymerase chain reaction with reverse transcription (RT-PCR)

OsGI, *Hd1* (*Se1*), *Hd3a* and *ubq* sequences were amplified by PCR with the primers listed in Supplementary Table 1. For the PCR experiments shown in Supplementary Fig. 2, conditions of 24, 28, 28 and 24 cycles were used for the amplification of *OsGI*, *Hd1* (*Se1*), *Hd3a* and *ubq*, respectively. For the PCR experiments shown in Fig. 3b, g and h, 19 and 17 cycles were used for the amplification of *Hd3a* and *ubq*, respectively. For real-time PCR used to obtain the results shown in Fig. 3i, quantitative analysis of gene expression was performed by SYBR Green PCR master mix (Applied Biosystems Japan, Tokyo, Japan) with the gene-specific primers listed in Supplementary Table 1. Data were collected using the ABI PRISM 7700 sequence detection system in accordance with the instruction manual.

RNAse protection assay

RNAse protection assays were performed with the RPA III kit (Ambion, Austin, Texas, USA). Antisense RNA probes for *OsGI*, *Hd1* (*Se1*) and *ubq* were synthesized with T3 or SP6 RNA polymerase (Promega, Madison, Wisconsin, USA). The relative radioactivity of the *Hd1* (*Se1*) probe was adjusted by increasing the concentration of [α -³²P]UTP and decreasing the concentration of unlabelled UTP in their reaction cocktails. A total of 30,000 c.p.m. of the purified RNA probe was added in a reaction buffer containing 20 μg of total RNA and annealed overnight with RNA; protected RNA probes were detected on 4% polyacrylamide gel with BAS 2000 (Fuji Photo Film, Tokyo, Japan). The gel image was analysed with MacBAS version 2.52 software to calibrate the signal intensity of each clone. The sizes of the protected RNAs were 83 bp for *OsGI*, 401 bp for *Hd1* (*Se1*) and 245 bp for *ubq*.

Transfection assays in rice protoplasts

Protoplasts (4 × 10⁷) were prepared from the rice Oc cell cultures, mixed with 30 μg of plasmid DNA containing *Hd3a::gus* and 10 μg of plasmid DNA containing *Ubiq::Hd1* (*Se1*) cDNA or *Ubiq::bar* and electroporated with a Gene Pulser (Bio-Rad). A 1.7-kb promoter region of the *Hd3a* gene was used to construct *Hd3a::gus*; it included an intron²³ to enhance *gus* expression. After incubation of protoplasts for 48 h at 30 °C in light, they were harvested and assayed for GUS activity.

Received 16 January 2002; accepted 3 March 2003; doi:10.1038/nature01549.

1. Thomas, B. & Vince-Prue, D. *Photoperiodism in Plants* (Academic, London, 1997).
2. Koornneef, M., Alonso-Blanco, C., Peeters, A. J. M. & Soppe, W. Genetic control of flowering time in *Arabidopsis*. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **49**, 345–370 (1998).
3. Mouradov, A., Cremer, F. & Coupland, G. Control of flowering time: Interacting pathways as a basis for diversity. *Plant Cell Suppl.* **14**, S111–S130 (2002).
4. Simpson, G. G. & Dean, C. *Arabidopsis*, the Rosetta stone of flowering time? *Science* **296**, 285–289 (2002).
5. Hayama, R., Izawa, T. & Shimamoto, K. Isolation of rice genes possibly involved in the photoperiodic control of flowering by a fluorescent differential display method. *Plant Cell Physiol.* **43**, 494–504 (2002).
6. Fowler, S. *et al.* GIGANTEA: A circadian clock-controlled gene that regulates photoperiodic flowering in *Arabidopsis* and encodes a protein with several possible membrane-spanning domains. *EMBO J.* **18**, 4679–4688 (1999).
7. Park, D. H. *et al.* Control of circadian rhythms and photoperiodic flowering by the *Arabidopsis* GIGANTEA gene. *Science* **285**, 1579–1582 (1999).
8. Yano, M. *et al.* *Hd1*, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the *Arabidopsis* flowering time gene *CONSTANS*. *Plant Cell* **12**, 2473–2484 (2001).
9. Putterill, J., Robson, E., Lee, K., Simon, R. & Coupland, G. The *CONSTANS* gene of *Arabidopsis* promotes flowering and encodes a protein showing similarities to zinc finger transcription factors. *Cell* **80**, 847–857 (1995).
10. Kojima, S. *et al.* *Hd3a*, a rice ortholog of the *Arabidopsis* *FT* gene, promotes transition to flowering downstream of *Hd1* under short-day conditions. *Plant Cell Physiol.* **43**, 1096–1105 (2002).
11. Kardailsky, I. *et al.* Activation tagging of the floral inducer *FT*. *Science* **286**, 1962–1965 (1999).
12. Kobayashi, Y., Kaya, H., Goto, K., Iwabuchi, M. & Araki, T. A pair of related genes with antagonistic roles in mediating flowering signals. *Science* **286**, 1960–1962 (1999).
13. Izawa, T., Oikawa, T., Tokutomi, S., Okuno, K. & Shimamoto, K. Phytochromes confer the photoperiodic control of flowering in rice (a short-day plant). *Plant J.* **22**, 391–399 (2000).

14. Araki, T. & Komeda, Y. Analysis of the role of the late-flowering locus, *GI*, in the flowering of *Arabidopsis thaliana*. *Plant J.* **3**, 231–239 (1993).
15. Samach, A. *et al.* Distinct roles of *CONSTANS* target genes in the reproductive development of *Arabidopsis*. *Science* **288**, 1613–1616 (2000).
16. Suarez-Lopez, P. *et al.* *CONSTANS* mediates between the circadian clock and the control of flowering in *Arabidopsis*. *Nature* **410**, 1116–1120 (2001).
17. Yanovsky, M. J. & Kay, S. Molecular basis of seasonal time measurement in *Arabidopsis*. *Nature* **419**, 308–312 (2002).
18. Izawa, T. *et al.* Phytochrome mediates the external light signal to repress *FT* orthologs in photoperiodic flowering of rice. *Genes Dev.* **16**, 2006–2020 (2002).
19. Martinez-Garcia, J. F., Virgos-Soler, A. & Prat, S. Control of photoperiod-regulated tuberization in potato by the *Arabidopsis* flowering-time gene *CONSTANS*. *Proc. Natl Acad. Sci. USA* **99**, 15211–15216 (2002).
20. Liu, J., Yu, J., McIntosh, L., Kende, H. & Zeevaert, J. A. D. Isolation of a *CONSTANS* ortholog from *Pharbitis nil* and its role in flowering. *Plant Physiol.* **125**, 1821–1830 (2001).
21. Christensen, A. H., Sharrock, R. A. & Quail, P. H. Maize polyubiquitin genes: Structure, thermal perturbation of expression and transcript splicing, and promoter activity following transfer to protoplasts by electroporation. *Plant Mol. Biol.* **18**, 675–689 (1992).
22. Hiei, Y., Ohta, S., Komari, T. & Kumashiro, T. Efficient transformation of rice (*Oryza sativa* L.) mediated by *Agrobacterium* and sequence analysis of the boundaries of the T-DNA. *Plant J.* **6**, 271–282 (1994).
23. Tanaka, A. *et al.* Enhancement of foreign gene expression by a dicot intron in rice but not tobacco is correlated with an increased level of mRNA and an efficient splicing of the intron. *Nucleic Acids Res.* **18**, 6767–6770 (1990).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements We thank Takeshi Izawa for help in the early part of this work, and members of the Plant Molecular Genetics Lab at NAIST for discussions. This research was supported by Grants-in-Aid for Scientific Research on Priority Areas of the Ministry of Education, Culture, Sports, Science, and Technology of Japan.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to K.S. (e-mail: simamoto@bs.aist-nara.ac.jp).

Amino-acid cycling drives nitrogen fixation in the legume–*Rhizobium* symbiosis

E. M. Lodwig^{*†}, A. H. F. Hosie^{*†}, A. Bourdès^{*}, K. Findlay[‡], D. Allaway^{*}, R. Karunakaran^{*}, J. A. Downie[‡] & P. S. Poole^{*}

^{*} School of Animal and Microbial Sciences, University of Reading, Whiteknights, PO Box 228, Reading RG6 6AJ, UK

[‡] John Innes Centre, Colney Lane, Norwich NR4 7UH, UK

[†] These authors contributed equally to this work

The biological reduction of atmospheric N₂ to ammonium (nitrogen fixation) provides about 65% of the biosphere's available nitrogen. Most of this ammonium is contributed by legume–rhizobia symbioses¹, which are initiated by the infection of legume hosts by bacteria (rhizobia), resulting in formation of root nodules. Within the nodules, rhizobia are found as bacteroids, which perform the nitrogen fixation: to do this, they obtain sources of carbon and energy from the plant, in the form of dicarboxylic acids^{2,3}. It has been thought that, in return, bacteroids simply provide the plant with ammonium. But here we show that a more complex amino-acid cycle is essential for symbiotic nitrogen fixation by *Rhizobium* in pea nodules. The plant provides amino acids to the bacteroids, enabling them to shut down their ammonium assimilation. In return, bacteroids act like plant organelles to cycle amino acids back to the plant for asparagine synthesis. The mutual dependence of this exchange prevents the symbiosis being dominated by the plant, and provides a selective pressure for the evolution of mutualism.

One of the most important aspects of legume–rhizobia symbioses is that bacteroids shut down ammonium assimilation. This could not be explained, although it is considered essential for the evolution of symbioses⁴. However, if plants provide an amino acid to bacteroids, then bacteroids need not assimilate ammonium⁵; instead, they would have to secrete ammonium to the plant to ensure their own amino-acid supply. It has been shown that when isolated peribacteroid units of peas are incubated with a dicarboxylate, addition of glutamate stimulates secretion of aspartate and alanine^{6,7}. Although this demonstrates the biochemical capacity for amino-acid cycling, the importance of this to nodule metabolism was hitherto unknown.

To determine whether amino-acid cycling occurs in pea bacteroids *in planta*, we mutated *aap* and *bra*, which both encode ABC-type broad specificity amino-acid transporters, in *Rhizobium leguminosarum* bv. *viciae*^{8–10}. Whereas single mutations in *aap* or *bra* reduced the uptake rates of all tested amino acids by 40–70% in free-living bacteria, the double mutant, RU1357, was almost totally blocked for the uptake of a broad range of amino acids, including glutamate, aspartate and leucine⁸. The double mutant grows on minimal medium, demonstrating that it synthesises amino acids. Whereas the growth of peas nodulated by either *aap* or *bra* single mutants, or by the wild type (A34), was indistinguishable, peas nodulated by RU1357 progressively yellowed (Fig. 1a). The marked reductions in shoot dry weight and shoot nitrogen content of peas nodulated by RU1357, as well as their increased nodule number and mass, are typical of plants unable to fix nitrogen (Fig. 2a–d). These effects can be attributed to the loss of amino-acid transport, as RU1357 carrying the *aap* operon on a stable plasmid (pRU1134) restored plants to normal growth.

Although nitrogen starvation of plants nodulated by RU1357 was very marked, the root nodules were pink, unlike the white nodules usually induced by classical non-fixing mutants, whereas wild-type nodules were deep red (Fig. 1b). The red colour is due to leghaemoglobin, a plant-made O₂-carrying protein, which is induced by RU1357 at a level intermediate between the wild type and a classical non-fixing mutant. The bacteroid yield from plants nodulated by RU1357 was 37% of that of A34. Acetylene and ¹⁵N₂ reduction

assays showed that plants nodulated by RU1357 fixed nitrogen at 32–55% of wild-type rates (Fig. 3a, c). When these rates are expressed on a bacteroid protein basis, the values for RU1357 are similar (¹⁵N₂ reduction) or even higher (acetylene reduction) than the wild type (Fig. 3b, d). Isolated bacteroids of RU1357 reduced ¹⁵N₂ at substantially higher rates than A34 (Fig. 3e, f). This shows that RU1357 bacteroids, blocked for amino-acid transport, can reduce N₂, but apparently the plant cannot acquire the resulting ammonium. This is in disagreement with our current models of both bacterial and plant-nodule metabolism.

We examined directly the fate of ¹⁵N₂ fixed in nodules by measuring xylem amide levels and ¹⁵N enrichment. The concentrations of xylem amides were extremely low in plants nodulated by RU1357 compared to A34 (Fig. 3g), with the concentration of asparagine for RU1357 only 11% of that for A34. Correcting the xylem amide concentration for bacteroid protein still leaves a low asparagine level for plants nodulated by RU1357 relative to the wild type (Fig. 3h). This is in contrast to ammonium production per unit bacteroid protein, which is similar between plants nodulated by A34 or RU1357 (Fig. 3d). As expected, there was high ¹⁵N enrichment of glutamine in the xylem sap of plants nodulated by A34, but also of RU1357 (Fig. 3i). Asparagine was also significantly enriched in ¹⁵N in plants nodulated by both A34 (0.82 atom% excess) and RU1357 (0.33 atom% excess), although predictably at lower levels than glutamine (Fig. 3i). These ¹⁵N enrichments show that plants retain the biochemical capacity to make amides. However, in the absence of amino-acid transport by bacteroids, plants cannot efficiently use the ammonium released, resulting in the extremely low amide concentration. An explanation is that bacteroids provide both ammonium and an amino acid such as aspartate for asparagine synthesis in the plant cytosol.

It is widely accepted that bacteroids accumulate dicarboxylates (fumarate, succinate or L-malate) via the dicarboxylic-acid transport system, and that these are oxidized to provide energy for N₂ reduction to ammonium^{11,12}. We propose a model in which glutamate (or a precursor to it) is transported into bacteroids, in addition to dicarboxylates (Fig. 4). The glutamate would enter via Aap/Bra, and act as a transamination donor to produce aspartate, or possibly amino acids such as alanine. These would be secreted to the plant enabling asparagine synthesis. Glutamate is the most probable donor amino acid, because it is highly abundant in nodules and is known to stimulate transamination of oxaloacetate and pyruvate in

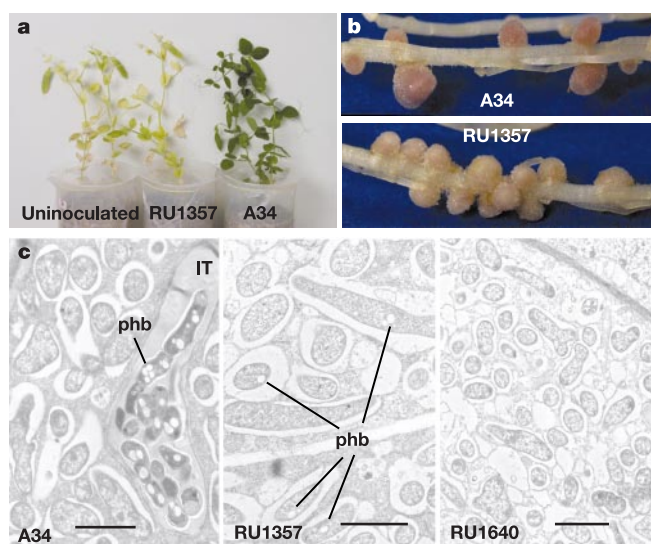


Figure 1 Effect of mutation of amino-acid uptake in *R. leguminosarum* on growth and nodulation of peas. **a**, Peas (*Pisum sativum* c.v. Avola) were left uninoculated, inoculated with *R. leguminosarum* strain RU1357 (*aap/bra*) or A34 (wild type) and grown for 40 d on nitrogen-free rooting solution. **b**, Three-week-old nodules of A34 and RU1357. **c**, Electron micrographs of plant cells infected with A34, RU1357 or RU1640 (*aat4*). Scale bars, 2 μ m. phb, polyhydroxybutyrate granule; it, infection thread.

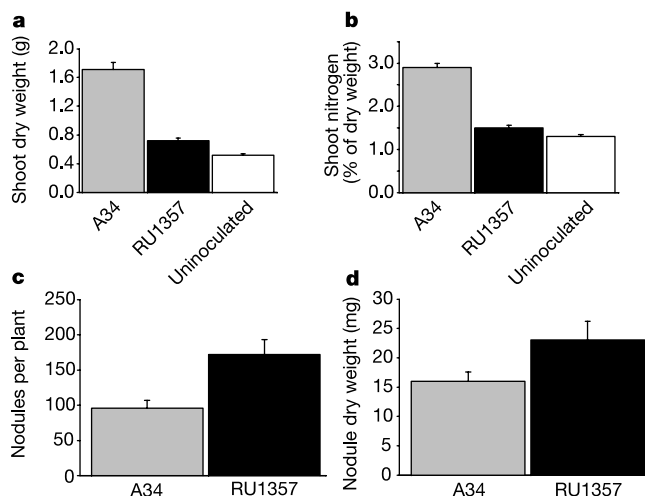


Figure 2 Growth and nodulation of peas infected by either wild type or a mutant unable to transport amino acids. **a**, Shoot dry weight ($n = 22$ plants). **b**, Shoot nitrogen as a percentage of dry weight ($n = 8$ plants). **c**, Nodule number per plant ($n \geq 7$ plants). **d**, Nodule dry weight per plant ($n \geq 7$ plants). All data shown $\pm$ s.e.m.

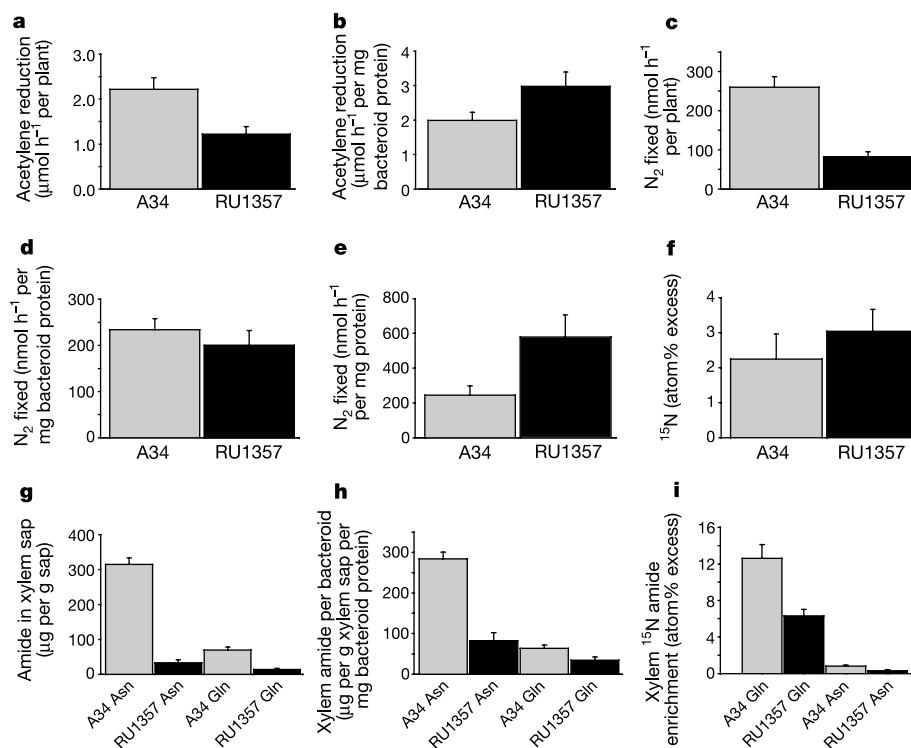


Figure 3 Nitrogen fixation and assimilation of peas infected by either wild type or a mutant unable to transport amino acids. **a**, Acetylene reduction by plants expressed per plant ($n \geq 12$ plants). **b**, Acetylene reduction by plants expressed per unit bacteroid protein ($n \geq 12$ plants). **c**, N_2 fixed by plants expressed per plant ($n \geq 9$ plants). **d**, N_2 fixed by plants expressed per unit bacteroid protein ($n \geq 9$ plants). **e**, Total N_2 reduced to

ammonium by isolated bacteroids ($n \geq 3$ assays). **f**, ^{15}N isotope enrichment of ammonium produced by isolated bacteroids ($n \geq 3$ assays). **g**, Xylem amide concentration per plant ($n = 9$ plants). **h**, Xylem amide concentration expressed per unit bacteroid protein ($n = 9$ plants). **i**, ^{15}N enrichment of pea xylem amides ($n = 9$ plants). All data shown $\pm$ s.e.m.

isolated pea peribacteroid units, leading to aspartate and alanine secretion^{6,7}. As Aap and Bra have a broad solute specificity, amino acids other than glutamate may be taken up from the plant. However, most bacterial amino-acid metabolism is channelled via glutamate, making little difference between direct uptake of glutamate or a precursor. Aspartate is the most likely secretion product, because glutamate stimulates its synthesis in isolated peribacteroid units, and a blockage in its secretion would dramatically reduce asparagine synthesis in the plant. However, the secretion of alanine by bacteroids could have a similar effect, because alanine, glutamate and aspartate can be interconverted in the plant by glutamate pyruvate transaminase and aspartate aminotransferase. We have previously shown that, apart from their role as uptake transporters, Aap and Bra also act as high-rate but low-affinity exporters, and therefore may also mediate the export of aspartate or other amino acids^{9,10}. Blocking amino-acid import, export or both will have similar effects.

There are two predictions of this model: first, bacteroid aspartate aminotransferase (Aat) activity should be essential for nitrogen fixation; and second, blocking amino-acid transport may cause bacteroids to become carbon saturated, because they can no longer remove dicarboxylic acids via transamination and amino-acid export.

Mutation of the principal aspartate aminotransferase gene (*aatA*) in *Sinorhizobium meliloti* blocks nitrogen fixation by alfalfa¹³. The reason for this was unknown, but it was considered unusual because an *aatA* mutant retained significant levels of Aat activity owing to other secondary aminotransferases. To determine if AatA is also essential for an effective symbiosis in pea, we generated an *aatA* mutant in A34 (RU1640). Strains A34, RU1640 and RU1640/pRU1133 (*aatA* cloned in pJP2) grown on glucose/ammonium minimal medium had Aat activities of $0.385 \pm 0.113 \mu\text{mol min}^{-1}$

per mg protein (mean $\pm$ s.e.m., $n = 4$), $0.003 \pm 0.003 \mu\text{mol min}^{-1}$ per mg protein ($n = 3$) and $1.343 \pm 0.427 \mu\text{mol min}^{-1}$ per mg protein ($n = 3$), respectively. Plants inoculated with RU1640 did not fix nitrogen, yellowing from 4 weeks and forming white nodules with no detectable acetylene reduction. Thus, as predicted, aspartate aminotransferase is needed in pea bacteroids for effective nitrogen fixation. Consistent with this, the activity of AatA in A34 bacteroids was high at $0.162 \pm 0.017 \mu\text{mol min}^{-1}$ per mg protein ($n = 3$).

Electron micrographs of nodule sections show that bacteroids mutated in *aatA* or *aap/bra* are fully developed, and are similar to both wild type and *dct* mutants^{11,12} (Fig. 1c). Light micrographs of nodules infected with RU1640 (*aatA*) or RU1357 (*aap/bra*) show a large accumulation of plant starch relative to A34. This accumulation is consistent with either no (RU1640), or lower (RU1357), symbiotic nitrogen fixation and therefore reduced carbon demand on the plant. Infection threads containing A34, RU1357 or RU1640 have undifferentiated bacteria containing electron-transparent polyhydroxybutyrate granules (Fig. 1c). Remarkably, only RU1357 bacteroids contain polyhydroxybutyrate granules. This is consistent with them using dicarboxylates inefficiently, because in the absence of imported glutamate they cannot make aspartate from oxaloacetate (Fig. 4). While A34 bacteroids would secrete aspartate, RU1357 presumably increases carbon flow from the dicarboxylate L-malate to pyruvate via malic enzyme. Pyruvate leads to polyhydroxybutyrate synthesis. The absence of polyhydroxybutyrate granules in *aatA* or *dctA* mutants is consistent with a blockage in dicarboxylate use.

The cessation of N_2 reduction in *aatA* or *dctA* mutants is more extreme than the phenotype of the amino-acid transport mutant RU1357. One possibility is that bacteroids mutated in *aatA* or *dctA* are metabolically disrupted by accumulated glutamate, which cannot be converted to aspartate. The *aap/bra* transport mutant would not accumulate glutamate^{8,9}. However, it is possible that *aatA*

Our model also predicts that the *in planta* effects of amino-acid transport mutations in bacteroids should be confined to ammonium assimilation in the nodule. To test this, pea plants nodulated by RU1357 were grown for three weeks on nitrogen-free medium, eliciting nitrogen starvation, and then either NaNO₃ (5 mM) or NH₄Cl (5 mM) was added. The plants rapidly greened, and growth recovered. Thus, although plants infected with strain RU1357 are defective in nitrogen assimilation in nodules, they retain the ability to assimilate nitrate or ammonium supplied to roots.

The permeability of the peribacteroid membrane to amino acids and organic acids is critical to the operation of any amino-acid cycle (Fig. 4). The pea peribacteroid membrane has an ammonium channel and a H^+ /aspartate export system that move solutes from

There are significant consequences of this model for the legume-*Rhizobium* symbiosis. As the plant provides bacteroids with amino acids, bacteroids can shut down ammonium assimilation. To obtain amino acids, bacteroids must secrete ammonium to the plant, enabling amino-acid synthesis. This provides a powerful selective pressure for the evolution of symbiosis, and suggests that the plant can regulate bacteroid dicarboxylate use by amino-acid supply. This could lead to the plant dominating the symbiosis. However, bacteroids act like plant organelles for aspartate synthesis, making the plant dependent on them. This provides a counter selection to plant dominance, and favours the evolution of mutualism. However, it is not a classical malate/aspartate shuttle, which transfers reductant but not carbon into mitochondria²⁰ and has been suggested might operate in bacteroids⁵. Bacteroids of RU1357 reduce N₂ to ammonium at specific rates similar to wild type, demonstrating they generate adequate reductant. The role of amino-acid cycling is to facilitate both dicarboxylate oxidation and ammonium assimilation into asparagine. The interaction between the symbiotic partners is far more complex than hitherto realized: one has evolved a complete metabolic dependence on the other. □

Strains A34 (wild type) and RU1357 (*ΔaapJQM::ΩSp braE::TnphoA*) were grown on acid minimal medium^{8,21}. An *aap/bra* double mutation (RU1722 *ΔaapJQM::ΩSp ΔbraEF::ΩTc*) was made in strain 3841. RU1722 has the *ΔaapJQM::ΩSp* mutation as in RU1357, but the *bra* system was mutated by deleting the *EcoRV* fragment that spans part of

braE and *braF*. Peas (*Pisum sativum* c.v. Avola and Winner) were grown in 2-l pots in vermiculite with nitrogen-free rooting solution²¹. Shoot dry weights were determined on 6-week-old plants by drying for 48 h at 70 °C, and total nitrogen determined by Dumas combustion. All other experiments on plants or bacteroids were determined with 3–4-week-old plants. Optical and electron-micrograph images were determined on sectioned nodules²². Total bacteroid protein per plant was determined by removing all nodules from plants, isolating bacteroids by Percoll purification²¹ and determining protein by the Lowry method. Acetylene reductions were determined on plants incubated in 95% air 5% acetylene for 1 h in 250-ml Schott bottles²¹. To measure ¹⁵N₂ reduction, plants were incubated in 80% air 20% ¹⁵N₂ (98 atom%, Isotec) for 1 h in 250-ml Schott bottles. The incorporation of ¹⁵N₂ was determined by separating roots and shoots, grinding them in 3–4 ml of HCl (0.1 M), drying overnight and measuring total nitrogen and ¹⁵N enrichment in a continuous flow isotope ratio mass spectrometer (TracerMat, Finnigan MAT)²¹. Synthesis of amino acids in pea xylem was measured in plants incubated as for total ¹⁵N₂ reduction, except that the incubation time was 30 min. Xylem sap was removed from the 3-cm stem section immediately below the first leaf²³. The concentrations and ¹⁵N enrichment of amino acids were determined with a Trace 2000 gas chromatograph (Finnigan) fitted with an AS 2000 autosampler (Finnigan) and interfaced to a Trace quadrupole mass spectrometer (Finnigan)²¹. Bacteroids were isolated in a microaerobic cabinet (MACS-MG-1000, Don Whately), under an atmosphere of 0.1% O₂, 1% CO₂ and 98.9% N₂, and ¹⁵N₂ reduction to ammonium was determined using a Roboprep combustion analyser (Europa Scientific) interfaced to a VG 622 isotope ratio mass spectrometer²¹ (VG).

The *R. leguminosarum aatA* gene was sequenced (EMBL AJ006709), mutated and recombined into strain A34 to create RU1640. The Gene Jumper kanamycin cassette (Invitrogen) in *aatA* is located between the 5-base-pair duplication of bases 409–413 (GGCAC), and its recombination into A34 to create strain RU1640 was confirmed by Southern blotting. The *aatA* gene was amplified by polymerase chain reaction with primers p383 (AGGTGACGCAAGCCTCTCCC) and p384 (ACCGTCACTCTGCCGTCACG), and cloned as an *XbaI/BamHI* fragment into the stable broad host range plasmid pJP2²⁴, creating pRU1133. The *aapJQMP* operon was cloned from pRU189¹⁰ as an *XbaI/HindIII* fragment into pJP2, creating pRU1134.

Received 29 November 2002; accepted 25 February 2003; doi:10.1038/nature01527.

- Newton, W. E. in *Nitrogen Fixation: From Molecules to Crop Productivity* (eds Pedrosa, F. O., Hungria, M., Yates, M. G. & Newton, W. E.) 3–8 (Kluwer, Dordrecht, 2000).
- Vance, C. P. in *Prokaryotic Nitrogen Fixation* (ed. Triplett, E.) 589–607 (Horizon Scientific, Wymondham, UK, 2000).
- Poole, P. S. & Allaway, D. Carbon and nitrogen metabolism in *Rhizobium*. *Adv. Microb. Physiol.* **43**, 117–163 (2000).
- Patriarca, E. J., Tate, R. & Iaccarino, M. Key role of bacterial NH₄⁺ metabolism in *Rhizobium*-plant symbiosis. *Microb. Mol. Biol. Rev.* **66**, 203–222 (2002).
- Kahn, M. L., Kraus, J. & Sommerville, J. E. in *Nitrogen Fixation Research Progress* (eds Evans, H. J., Bottomley, P. J. & Newton, W. E.) 193–199 (Martinus Nijhoff, Dordrecht, 1985).
- Rosendahl, L., Dilworth, M. J. & Glenn, A. R. Exchange of metabolites across the peribacteroid membrane in pea root nodules. *J. Plant Physiol.* **139**, 635–638 (1992).
- Appels, M. A. & Haaker, H. Glutamate oxaloacetate transaminase in pea root nodules—participation in a malate/aspartate shuttle between plant and bacteroid. *Plant Physiol.* **95**, 740–747 (1991).
- Hosie, A. H. F., Allaway, D., Dunsky, H. A., Galloway, C. S. & Poole, P. S. *Rhizobium leguminosarum* has a second general amino acid permease with unusually broad substrate specificity and high similarity to branched-chain amino acid transporters (Bra/LIV) of the ABC family. *J. Bacteriol.* **184**, 4071–4080 (2002).
- Hosie, A. H. F. *et al.* Solute-binding protein-dependent ABC transporters are responsible for solute efflux in addition to solute uptake. *Mol. Microbiol.* **40**, 1449–1459 (2001).
- Walshaw, D. L. & Poole, P. S. The general L-amino acid permease of *Rhizobium leguminosarum* is an ABC uptake system that influences efflux of solutes. *Mol. Microbiol.* **21**, 1239–1252 (1996).
- Ronson, C. W., Lyttleton, P. & Robertson, J. G. C₄-dicarboxylate transport mutants of *Rhizobium trifolii* form ineffective nodules on *Trifolium repens*. *Proc. Natl Acad. Sci. USA* **78**, 4284–4288 (1981).
- Finan, T. M., Wood, J. M. & Jordan, D. C. Symbiotic properties of C₄-dicarboxylic acid transport mutants of *Rhizobium leguminosarum*. *J. Bacteriol.* **154**, 1403–1413 (1983).
- Watson, R. J. & Rastogi, V. K. Cloning and nucleotide sequencing of *Rhizobium meliloti* aminotransferase genes: an aspartate aminotransferase required for symbiotic nitrogen fixation is atypical. *J. Bacteriol.* **175**, 1919–1928 (1993).
- Craig, J. *et al.* Mutations at the *rug4* locus alter the carbon and nitrogen metabolism of pea plants through an effect on sucrose synthase. *Plant J.* **17**, 353–362 (1999).
- Saalbach, G., Erik, P. & Wienkoop, S. Characterisation by proteomics of peribacteroid space and peribacteroid membrane preparations from pea (*Pisum sativum*) symbiosomes. *Proteomics* **2**, 325–337 (2002).
- Rosendahl, L., Mouritzen, P. & Rudbeck, A. Nitrogen transfer in the interface between the symbionts in pea root nodules. *Plant Soil* **230**, 31–37 (2001).
- Mouritzen, P. & Rosendahl, L. Identification of a transport mechanism for NH₄⁺ in the symbiosome membrane of pea root nodules. *Plant Physiol.* **115**, 519–526 (1997).
- Udvardi, M. K., Price, G. D., Gresshoff, P. M. & Day, D. A. A dicarboxylate transporter on the peribacteroid membrane of soybean nodules. *FEBS Lett.* **231**, 36–40 (1988).
- Udvardi, M. K., Salom, C. L. & Day, D. A. Transport of L-glutamate across the bacteroid membrane but not the peribacteroid membrane from soybean root nodules. *Mol. Plant-Microbe Interact.* **1**, 250–254 (1988).
- Indiveri, C., Krämer, R. & Palmieri, F. Reconstitution of the malate/aspartate shuttle from mitochondria. *J. Biol. Chem.* **262**, 15979–15983 (1987).
- Allaway, D. *et al.* Identification of alanine dehydrogenase and its role in mixed secretion of ammonium and alanine by pea bacteroids. *Mol. Microbiol.* **36**, 508–515 (2000).
- Beringer, J. E., Johnston, A. W. B. & Wells, B. The isolation of conditional ineffective mutants of *Rhizobium leguminosarum*. *J. Gen. Microbiol.* **98**, 339–343 (1977).
- Parsons, R. & Baker, A. Cycling of amino compounds in symbiotic lupin. *J. Exp. Bot.* **47**, 421–429 (1996).

- Prell, J., Boesten, B., Poole, P. & Priefer, U. B. The *Rhizobium leguminosarum* bv. *viciae* VF39 gamma aminobutyrate (GABA) aminotransferase gene (*gabT*) is induced by GABA and highly expressed in bacteroids. *Microbiology* **148**, 615–623 (2002).

Acknowledgements We thank the BBSRC for supporting this research. We thank A. Hepburn and M. Heaps for GC-MS and nitrogen analysis and A. East for manuscript preparation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.S.P. (e-mail: p.s.poole@reading.ac.uk).

Transcription-targeted DNA deamination by the AID antibody diversification enzyme

Jayanta Chaudhuri, Ming Tian, Chan Khuong, Katrin Chua, Eric Pinaud & Frederick W. Alt

Howard Hughes Medical Institute, The Children's Hospital, The Center for Blood Research, and Department of Genetics, Harvard University Medical School, Boston, Massachusetts 02115, USA

Activation-induced cytidine deaminase (AID), which is specific to B lymphocytes, is required for class switch recombination (CSR)—a process mediating isotype switching of immunoglobulin—and somatic hypermutation—the introduction of many point mutations into the immunoglobulin variable region genes^{1,2}. It has been suggested that AID may function as an RNA-editing enzyme³ or as a cytidine deaminase on DNA^{4,5}. However, the precise enzymatic activity of AID has not been assessed in previous studies. Similarly, although transcription of the target immunoglobulin locus sequences is required for both CSR and somatic hypermutation, the precise role of transcription has remained speculative^{6–9}. Here we use two different assays to demonstrate that AID can deaminate specifically cytidines on single-stranded (ss)DNA but not double-stranded (ds)DNA substrates *in vitro*. However, dsDNA can be deaminated by AID *in vitro* when the reaction is coupled to transcription. Moreover, a synthetic dsDNA sequence, which targets CSR *in vivo* in a manner dependent on transcriptional orientation¹⁰, was deaminated by AID *in vitro* with the same transcriptional-orientation-dependence as observed for endogenous CSR. We conclude that transcription targets the DNA deamination activity of AID to dsDNA by generating secondary structures that provide ssDNA substrates.

Immunoglobulin heavy (IgH) and light (IgL) chain variable region exons are assembled in developing B-lineage cells by recombination of V, D and J segments—V(D)J recombination¹¹. After antigen-dependent activation, mature B cells undergo additional immunoglobulin locus alterations, namely CSR and somatic hypermutation^{6–9}. CSR changes the expressed IgH constant region exons (C_H) from C_μ to a downstream C_H (for example, C_γ, C_ε, C_α), allowing association of the variable region with different C_H effector functions^{6,7}. CSR occurs between 1–12-kilobase (kb) repetitive switch (S) region sequences located 5' of each C_H. S regions are required for efficient CSR^{10,12}, as is transcription through an S region before CSR^{6,13}. In this regard, a 1-kb synthetic sequence that generates a G-rich transcript, but that lacks S region motifs, effects CSR in a transcriptional-orientation-dependent manner in place of S_{γ1} *in vivo*¹⁰. This finding suggested that ssDNA that is stabilized by transcription-dependent higher-order structures might be a

substrate for CSR-initiating factors¹⁰. Somatic hypermutation introduces mutations at high frequency into IgH and IgL variable region exons, allowing selection of B cells that secrete higher-affinity immunoglobulin. Somatic hypermutation is preferentially targeted to hypermutation domains consisting of RGYW motifs (where R is A or G, Y is C or T, and W is A or T) and also has been linked to target sequence transcription⁷⁻⁹.

CSR and somatic hypermutation require AID^{1,2}, and AID expression can induce CSR and somatic hypermutation in substrates in lymphoid and non-lymphoid cells, suggesting that it is the only required B-cell-specific activity^{14,15}. AID has structural similarity to the APOBEC1 RNA-editing cytidine deaminase, indicating, along with other findings, that AID might function by editing messenger RNAs that then encode proteins essential for CSR and somatic hypermutation¹⁻³. In contrast, other studies suggest that AID functions as a DNA-specific cytidine deaminase, leading to a model proposing that AID-mediated DNA deamination generates G•U mismatches in S regions or somatic hypermutation hotspots that can be differentially resolved to effect somatic hypermutation or CSR^{4,5,16}. However, in the absence of direct evidence for AID enzymatic activity, the actual enzymatic function of AID has remained unclear³. Moreover, no clear answer has been provided to the long-standing question of how transcription can target a specific recombination event.

To assay DNA cytidine deaminase activity (CDA) *in vitro*, we designed a model synthetic DNA substrate (RGYW substrate)

comprising multimers of the RGYW somatic hypermutation hot-spot sequence. A DNA cytidine deaminase enzyme will convert one or more cytidines in the substrate to uridine, which can then be excised from the DNA by either endogenous or supplemented recombinant uracil DNA glycosylase (UNG and UDG, respectively). If the cytidines in the substrate are [³H]-labelled, cytidine deamination followed by UNG activity will release [³H]-labelled uracil. The released [³H]-labelled uracil can be measured as a trichloroacetic acid (TCA) soluble count that would be proportional directly to DNA CDA. We assayed the ability of activated splenic B-cell extracts containing AID protein (Fig. 1a) to catalyse DNA CDA, and found no stimulation of CDA on native double-stranded RGYW substrate compared with that observed with extracts from AID-deficient B cells (Fig. 1b). If, however, the RGYW substrate was first denatured to generate ssDNA, DNA CDA was readily and reproducibly observed in wild-type but not AID-deficient extracts over a wide range of protein concentrations (Fig. 1b, c). We also assayed DNA CDA in extracts from a human embryonic kidney cell line (293) that expressed AID from a transfected construct, and found activity (on ssDNA) in AID-expressing extracts but not mock-transfected controls (Fig. 1b). As a control, we expressed a mutant AID (H56R/E58Q) that abrogates CDA on free nucleotides¹⁷, and found no detectable DNA CDA (Fig. 1a, b). Finally, to control for extract integrity, we assayed endogenous UNG activity and observed comparable levels in all extracts (Fig. 1e).

To establish that the release of [³H] measured actual cytidine

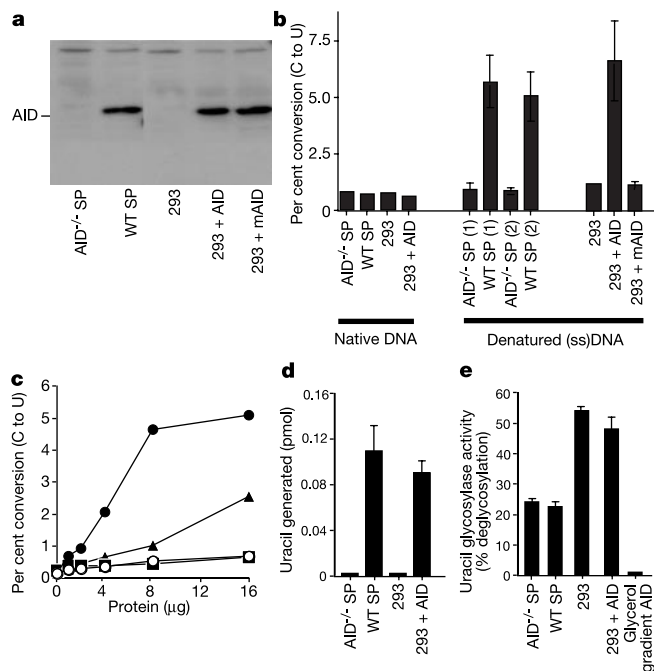


Figure 1 DNA cytidine deaminase activity in AID-expressing cells. **a**, Extracts (8 μg) derived from two independent experiments of the activation of wild-type and AID-deficient splenic B cells (WT SP and AID^{-/-} SP, respectively) or from 293 cells transiently transfected with vector alone (293), wild-type AID complementary DNA (293 + AID) or mutant cDNA (293 + mAID), analysed by western blot using anti-AID antibodies. **b**, Assay of cell extracts (10 μg) for DNA CDA on native DNA and ssDNA. **c**, Wild-type (filled circles) and AID^{-/-} (open circles) B-cell extracts assayed for ssDNA CDA activity. Wild-type extracts incubated with 10 μM (triangles) or 20 μM (squares) tetrahydrouridine were also assayed. **d**, The amount of uracil base generated by cytidine deamination of ssDNA substrate (2 pmol) was directly determined by thin-layer chromatography. **e**, Measurement of the activity of uracil deglycosylases in cell extracts (4 μg) determined by the release of [³H]-labelled uracil from single-stranded [³H]dU-DNA.

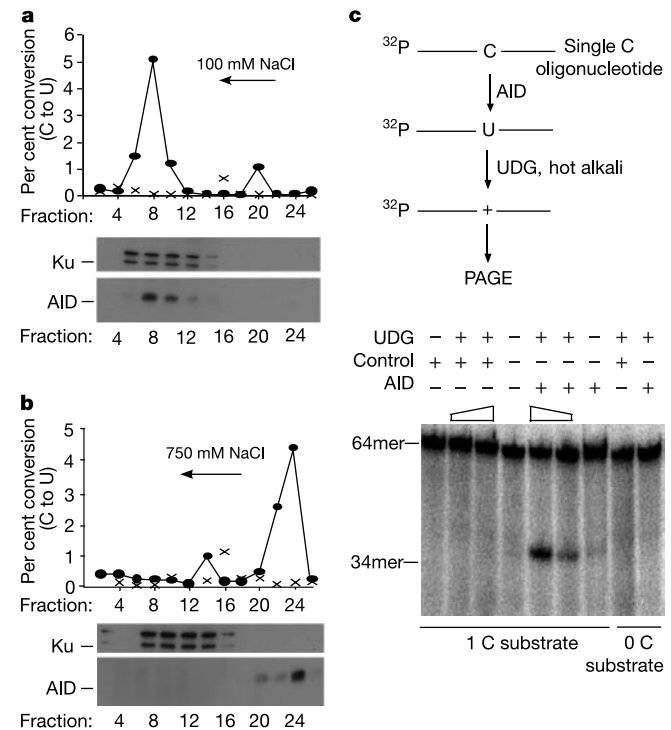


Figure 2 ssDNA cytidine deaminase activity co-sediments with AID in glycerol gradients. **a**, **b**, Extracts (400 μg) derived from vector-transfected or AID cDNA-transfected 293 cells were sedimented through glycerol gradients at either 100 mM (**a**) or 750 mM NaCl (**b**). Alternate fractions were then assayed for ssDNA CDA and analysed by western blot using anti-AID and anti-Ku antibodies. Filled circles and crosses represent activity in AID cDNA and vector-transfected extracts, respectively. The arrow indicates the direction of sedimentation. **c**, Glycerol-gradient-purified AID (2 μg or 4 μg) or similar fractions (control) derived from vector-transfected cells were incubated with a [5'-³²P] 64-base oligonucleotide (64mer) containing either one or no cytidine residue. The reaction products were deglycosylated by UDG, treated with hot alkali and then analysed by polyacrylamide gel electrophoresis (PAGE) followed by autoradiography.

deamination, we assayed a substrate in which the [^3H]-label was in guanosine instead of cytidine, and observed background levels of [^3H]-release for both wild-type and AID-deficient spleen extracts (not shown). Furthermore, a cytidine-deaminase-specific inhibitor, tetrahydrouridine¹⁸, markedly inhibited the activity of wild-type splenic B-cell extracts (Fig. 1c). Finally, we analysed reaction products by thin-layer chromatography and demonstrated that [^3H]-labelled uracil base and not [^3H]-labelled deoxyuridine is generated (Fig. 1d), ruling out the possibility that the assay reflected deamination of deoxycytidine nucleotides generated through non-specific nuclease activity. We conclude that our [^3H]-release assay measures specific conversion of cytidines to uridines on DNA.

The above experiments correlate ssDNA CDA with the presence of AID, but do not prove that the DNA cytidine deaminase was AID, as AID might edit a mRNA that encodes a DNA dC-deaminase. To address this issue, we subjected partially purified AID-expressing 293 cell extracts to glycerol gradient centrifugation, and assayed fractions for ssDNA CDA. Under low-salt conditions (100 mM NaCl), most of the DNA CDA and AID protein (detected by western blotting) sedimented as a complex with a relative molecular mass of over 500,000 (500K; peak of AID protein and DNA CDA in fraction 8, Fig. 2a); under high-salt conditions (750 mM NaCl), however, the DNA CDA and AID protein sedimented in the 30–60K fractions (peak of AID protein and DNA CDA in fraction 24, Fig. 2b). The high-salt fractions enriched for AID were pooled and used in an alternative DNA CDA assay. For this we designed an oligonucleotide substrate that contained only one cytidine residue amid RGYW motifs, and assessed the ability of AID-enriched fractions to target specifically the single cytidine residue. The abasic site created by this reaction then could be revealed by alkaline cleavage and gel electrophoresis. Specific cleavage at the cytidine residue was observed in an AID-dependent reaction and was dependent on addition of recombinant UDG (Fig. 2c), as the glycerol gradient fraction assayed did not contain UNG activity (Fig. 1e). Specificity of the site-specific deamination reaction was confirmed by the

absence of cleaved fragment from a substrate that lacked a cytidine residue (Fig. 2c).

Additional direct evidence that AID is a ssDNA-specific cytidine deaminase was obtained from immunodepletion using AID-specific antibodies. Complete AID immunodepletion from activated wild-type splenic extracts required two steps, with the supernatant from the first depletion having reduced AID levels, whereas the second completely lacked AID (Fig. 3a). The partially depleted extract exhibited decreased ssDNA-specific dC-deamination activity and the completely depleted extract had background activity (Fig. 3b, left panel). Moreover, activity in the completely depleted extract was restored when supplemented with AID immunoprecipitate from overexpressing 293 cells, but not with those of mock transfectants or a Myc-tagged immunoprecipitate from AID transfectants (Fig. 3b, middle panel). Furthermore, re-suspended AID immunoprecipitates displayed substantial DNA CDA that was dependent on the addition of recombinant UDG, confirming that the reaction proceeds through conversion of cytidine to uridine residues in DNA (Fig. 3b, right panel). No DNA CDA was observed in control immunoprecipitates (Fig. 3, right panel). These results indicate that AID is a ssDNA-specific cytidine deaminase, although they do not exclude additional cofactors, as immunoprecipitates and gradient fractions contained other proteins (data not shown).

The above results show that AID is a ssDNA-specific dC-deaminase with little or no activity on dsDNA. In this regard, the amino group of cytidine in dsDNA may not be accessible to the deaminase,

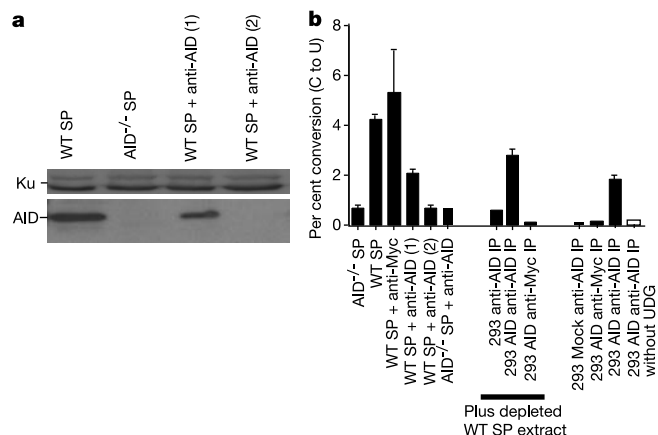


Figure 3 Immunoprecipitated AID contains ssDNA cytidine deaminase activity. **a**, Activated splenic B-cell extracts from wild-type or AID^{-/-} mice were subjected to two rounds of immunodepletion (1 and 2) with anti-AID antibodies and excess protein G Sepharose. Specific depletion of AID protein was analysed in western blots using anti-AID and anti-Ku antibodies. **b**, Immunodepleted supernatants assayed for ssDNA CDA (left panel). The middle panel shows AID immunoprecipitates from vector-transfected (293 anti-AID IP) or AID cDNA-transfected 293 cells (293 AID anti-Myc IP) added to AID-depleted supernatant (number 2) to reconstitute ssDNA CDA activity. In the right panel, the immunoprecipitates were assayed directly for ssDNA CDA. Unless otherwise indicated (open bar), all reactions contained recombinant bacterial UDG. Anti-Myc antibody was used as a nonspecific antibody control in the immunodepletion and reconstitution experiments.

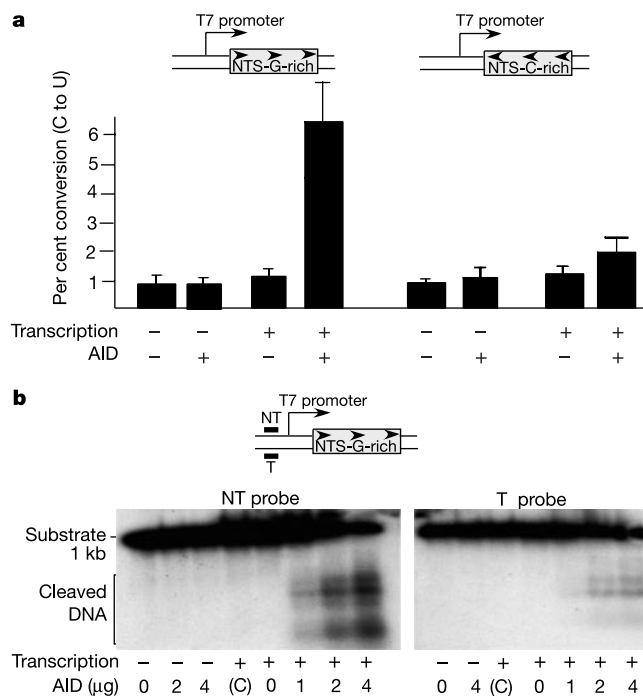


Figure 4 Transcription-coupled dsDNA deamination by AID. **a**, [^3H]cytidine-labelled G-rich or C-rich synthetic DNA substrates (NTS-G-rich and NTS-C-rich, respectively) were transcribed *in vitro* with T7 RNA polymerase. Hydroxyapatite (HAP)-purified AID (2 μg , (+) AID lanes) or HAP-purified control fractions (2 μg , (-) AID lanes) were added to the transcription reaction and deamination was assayed by measuring [^3H]-labelled uracil release. **b**, The NTS-G-rich substrate was subjected to coupled transcription–deamination, followed by deglycosylation and alkaline hydrolysis. The DNA recovered from the reaction was denatured and analysed by Southern hybridization using oligonucleotide probes specific to the template (T) or non-template (NT) strand. (C), reactions without AID but with 5 μg HAP-purified mock-transfected extracts; NTS, non-transcribed strand. Arrows indicate the orientation of the G-rich strand with respect to transcription.

as it is hydrogen-bonded to guanine. CSR and somatic hypermutation both require transcription through target sequences⁷ that, at least for S regions *in vitro*, can generate secondary structures^{19,20} including stable R loops^{10,21,22} and stem loops²³ that could provide ssDNA targets. Mammalian S regions are G-rich on their physiological non-template strand (NTS-G-rich) and form stable R loops when transcribed in this direction *in vitro*^{10,21}. R-loop structures are also formed when a synthetic NTS-G-rich substrate, but not its inverted NTS-C-rich configuration, is transcribed *in vitro*¹⁰. Moreover, the potential to form such structures correlates with the ability of this synthetic sequence in the NTS-G-rich, but not the inverted NTS-C-rich, orientation to support CSR in place of Sy1 *in vivo*¹⁰. Therefore, to test the potential role of transcription, we transcribed the NTS-G-rich and NTS-C-rich synthetic sequences *in vitro* using a T7 promoter/T7 RNA polymerase system, and evaluated the ability of AID to mediate DNA CDA on the transcribed duplex template.

We observed significant DNA CDA when the NTS-G-rich substrate was transcribed; however, we observed modest DNA CDA when the NTS-C-rich DNA was transcribed (Fig. 4a). Similar results were obtained with our DNA cleavage assay (not shown). Furthermore, the same orientation-dependence on deamination was observed when DNA CDA was analysed after removal of T7 RNA polymerase following transcription (data not shown). This indicates that DNA deamination need not be co-transcriptional and that AID can deaminate stable R loops subsequent to their generation by transcription. Finally, to determine whether one or both strands are deaminated, we assessed relative deamination of the ssDNA in the R loop (non-template strand) compared with the template strand of the NTS-G-rich substrate. For this purpose we performed deglycosylation and alkaline hydrolysis of reaction products from coupled transcription–deamination reactions followed by denaturation and analysis by Southern hybridization with strand-specific oligonucleotide probes. Consistent with DNA CDA of AID, we found a clear AID-dependent decrease in the amount of the non-template strand with the concomitant appearance of smaller fragments—the template strand also was deaminated, although at much lower levels (Fig. 4b). We conclude that transcription can target the DNA CDA of AID to the synthetic dsDNA sequence in an orientation-dependent fashion, with deamination preferentially, but not exclusively, targeted to the displaced single strand in the R loop. We were unable to detect AID-mediated deamination of RNA transcribed from the NTS-G-rich substrate (generating an RNA corresponding to the displaced G-rich strand) (data not shown), indicating that this transcript is not a preferred substrate for AID. However, this finding does not exclude AID acting specifically on some unknown RNA motif.

Our findings suggest a model in which transcription through the S region, along with its unique sequence, generates secondary structures during CSR that reveal ssDNA stretches, which serve as AID DNA cytidine deamination substrates. The resulting dU-DNA could then be processed by UNG and AP endonuclease to generate single-strand nicks that when occurring on both S region strands could lead to double-strand breaks, serving as a CSR intermediates^{4,5}. This overall model provides a satisfactory explanation for the essential role of unique S region structures and of transcription in CSR. Furthermore, whereas AID is preferentially targeted to the looped-out substrate strand, it acts at low levels on the other strand as well. In fact, we have demonstrated that regions of ssDNA exist on both strands at the duplex DNA–R-loop transition area^{10,21}, which if used by such a mechanism might produce observed biases in location of S region junctions within fused upstream and downstream S regions²⁴. Also, additional structures including stem loops²³, possibly potentiated by other factors²⁵, might be formed on transcribed endogenous S regions owing to their motif structure, to provide alternative ssDNA substrates for AID.

Our results may also be relevant to the finding that somatic hypermutation occurs as a ssDNA event in the human BL2 line²⁶.

For somatic hypermutation, however, there may be separate factors that generate AID-accessible ssDNA substrates, as we have not, so far, found AID-mediated deamination on the RGYW dsDNA substrate, even when actively transcribed (data not shown). In bacteria, sequence-specific transcriptional pausing exposes the non-template strand and renders it accessible to additional proteins²⁷. During somatic hypermutation, RNA polymerase II holoenzyme (or pol I) might generate ssDNA substrates when it pauses during elongation²⁸, either due to transient ssDNA in the transcription bubble or to the activity of additional unidentified factors. In the latter context, it is notable that only a small subset of actively transcribed genes undergoes somatic hypermutation⁸. Finally, transcription need not be coupled actively to AID deamination in CSR, as we find that stable R loops are targets for AID in the absence of active transcription, and S regions form R loops *in vivo*²⁹. Thus, even though both processes require transcription, the nature of the mechanism by which the substrate ssDNA for AID is generated may mechanistically distinguish CSR from somatic hypermutation.

Note added in proof: We have recently been made aware of a paper by Ramiro *et al.*³¹ that describes findings relevant to those reported here. □

Methods

DNA substrates

To construct the RGYW somatic hypermutation substrate, a 60-base oligonucleotide containing RGYW repeats (5′-AGCTGGCAGGCTAGCAAGTTGGTGGCAAGCAGGTAAGCAGG CAAGCTGGCTGAATTC-3′) was annealed to its complementary strand, trimerized using T4 DNA ligase and then cloned into pBluescript. To generate [³H]-labelled DNA substrate, the RGYW sequence was amplified by polymerase chain reaction (PCR) using M13-forward and M13-reverse primers and [³H]dCTP (Amersham) instead of dCTP. We used the purified PCR product as the substrate in the dsDNA CDA reaction. To generate ssDNA substrate, the dsDNA was heated for 30 min at 100 °C and then chilled on ice. The cloning of the G-rich and C-rich DNA substrates has been described¹⁰. The [³H]-labelled G-rich and C-rich substrates with the upstream T7 promoter were generated by PCR using primers derived from the vector sequence.

Extract preparation and purification of AID

Purified B cells from wild-type and AID-deficient mice were stimulated with anti-CD40 antibody and interleukin-4 as described¹⁰. After 4 days in culture, the cells were collected and lysed in buffer A (20 mM Tris-Cl, pH 7.5, 10 mM NaCl, 1 mM dithiothreitol (DTT) and 10 μM ZnCl₂) by dounce homogenization. Whole-cell extracts were then prepared by adding an equal volume of buffer B (50 mM Tris-Cl, pH 7.5, 1 M NaCl, 2 mM DTT, 0.5% NP-40, 20 μM ZnCl₂ and 20% glycerol), dounce homogenization and centrifugation. The lysate was dialysed for 4 h against buffer A plus 300 mM NaCl (buffer A-300) and then loaded onto a DEAE-cellulose column equilibrated with buffer A-300. The flow through was collected, dialysed against buffer A-75 and frozen. This fraction constitutes the crude extracts with which DNA CDA assays were performed. Transfected 293 cells were collected 60–72 h after transfection and extracts were prepared as above. For glycerol gradient centrifugation, the dialysed DEAE-cellulose fraction was first concentrated through Centricon-10 filtration and then layered onto a 15–40% continuous glycerol gradient containing buffer A-100 or buffer A-750. The protein extracts were sedimented in a SW41 rotor for 20 h at 21,000g. Fractions (400 μl) were collected from the bottom of the tube, dialysed against buffer A-100 and analysed. For further purification, glycerol gradient fractions containing AID and the corresponding fractions from mock-transfected extracts were pooled and loaded onto a hydroxyapatite column equilibrated with 10 mM Na-PO₄ buffer (pH 7.2) and then eluted with 50 mM Na-PO₄ (pH 7.2). The eluted proteins were dialysed against buffer A-100 containing 55% glycerol and then stored at –20 °C.

AID antibodies and immunoprecipitation

A peptide derived from the carboxy terminus of AID (EVDDLRLDAFRMLGF)¹⁸ was coupled to keyhole limpet haemocyanin and injected into rabbits as antigen. Antibodies were affinity purified against the peptide and used in western blot assays and immunoprecipitations. For the immunoprecipitations, extracts were incubated with anti-AID or anti-Myc antibodies and an excess of protein G Sepharose for 10 h in a buffer containing 50 mM Tris-Cl, pH 7.5, 350 mM NaCl and 0.1% Tween-20. The immunoprecipitate was washed several times with the same buffer and finally resuspended in buffer A-100. The immunoprecipitate was assayed directly for DNA CDA activity. The immunodepleted supernatant was dialysed briefly against buffer A-100 and assayed.

DNA cytidine deamination assay

For the uracil release assay, the [³H]-labelled DNA (1–2 pmol, specific activity 20,000–30,000 c.p.m. per pmol) substrate was incubated with 1–16 μg of protein preparation and recombinant bacterial UDG (NEB Biolabs) for 1 h at 37 °C in a buffer containing 25 mM Tris-Cl, pH 7.5, 100 mM NaCl, 1 mM DTT and 10 μM ZnCl₂. The reaction was stopped by the addition of TCA to 1%, centrifuged, and the radioactivity in the supernatant

determined in a liquid scintillation counter. DNA CDA activity was expressed as a percentage of TCA soluble to total [^3H] counts in the input DNA substrate.

For the DNA nicking activity, a 64-base oligonucleotide (5'-GGAATT GAGTTGGTAGGGTAGTAAGTTGGTTGGCAAGGAGGTAAGTAGGGAAGATGGAT GAAAT-3') was labelled with [γ - ^{32}P]ATP and T4 polynucleotide kinase. This was incubated with AID (glycerol gradient fractions), deglycosylated and then analysed in a standard 'nicking assay' as described for UDG³⁰.

Transcription of the G-rich and C-rich DNA was performed as described¹⁰. To determine which strand of the G-rich DNA is deaminated by AID, after incubation with AID the DNA was deglycosylated, deproteinated with proteinase K, treated with hot alkali and then precipitated. Denaturation and Southern analysis of the DNA was as described previously²¹.

Uracil DNA glycosylase assay

The RGYW DNA was labelled by PCR with [^3H]-labelled uridine, purified and then incubated with extracts at 37 °C for 1 h in buffer containing 20 mM Tris-Cl, pH 7.5, 50 mM NaCl and 1 mM DTT. The release of uracil was measured as TCA soluble count, the same as that described for DNA CDA.

Received 23 February; accepted 24 March 2003; doi:10.1038/nature01574.

Published online 9 April 2003.

1. Muramatsu, M. *et al.* Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* **102**, 553–563 (2000).
2. Revy, P. *et al.* Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). *Cell* **102**, 565–575 (2000).
3. Doi, T., Kinoshita, K., Ikegawa, M., Muramatsu, M. & Honjo, T. *De novo* protein synthesis is required for the activation-induced cytidine deaminase function in class-switch recombination. *Proc. Natl Acad. Sci. USA* **18**, 2634–2638 (2003).
4. Petersen-Mahrt, S. K., Harris, R. S. & Neuberger, M. S. AID mutates *E. coli* suggesting a DNA deamination mechanism for antibody diversification. *Nature* **418**, 99–103 (2002).
5. Rada, C. *et al.* Immunoglobulin isotype switching is inhibited and somatic hypermutation perturbed in UNG-deficient mice. *Curr. Biol.* **12**, 1748–1755 (2002).
6. Manis, J. P., Tian, M. & Alt, F. W. Mechanism and control of class-switch recombination. *Trends Immunol.* **23**, 31–39 (2002).
7. Honjo, T., Kinoshita, K. & Muramatsu, M. Molecular mechanism of class switch recombination: linkage with somatic hypermutation. *Annu. Rev. Immunol.* **20**, 165–196 (2002).
8. Martin, A. & Scharff, M. D. AID and mismatch repair in antibody diversification. *Nature Rev. Immunol.* **2**, 605–614 (2002).
9. Papavasiliou, F. N. & Schatz, D. G. Somatic hypermutation of immunoglobulin genes: merging mechanisms for genetic diversity. *Cell* **109** (suppl.), 35–44 (2002).
10. Shinkura, R. *et al.* The effect of transcriptional orientation on endogenous switch region function. *Nature Immunol.* (in the press).
11. Bassing, C. H., Swat, W. & Alt, F. W. The mechanism and regulation of chromosomal V(D)J recombination. *Cell* **109** (suppl.), 45–55 (2002).
12. Luby, T. M., Schrader, C. E., Stavnezer, J. & Selsing, E. The mu switch region tandem repeats are important, but not required, for antibody class switch recombination. *J. Exp. Med.* **193**, 159–168 (2001).
13. Stavnezer, J. Molecular processes that regulate class switching. *Curr. Top. Microbiol. Immunol.* **245**, 127–168 (2000).
14. Martin, A. *et al.* Activation-induced cytidine deaminase turns on somatic hypermutation in hybridomas. *Nature* **415**, 802–806 (2002).
15. Yoshikawa, K. *et al.* AID enzyme-induced hypermutation in an actively transcribed gene in fibroblasts. *Science* **296**, 2033–2036 (2002).
16. Di Noia, J. & Neuberger, M. S. Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase. *Nature* **419**, 43–48 (2002).
17. Papavasiliou, F. N. & Schatz, D. G. The activation-induced deaminase functions in a postcleavage step of the somatic hypermutation process. *J. Exp. Med.* **195**, 1193–1198 (2002).
18. Muramatsu, M. *et al.* Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. *J. Biol. Chem.* **274**, 18470–18476 (1999).
19. Reaban, M. E. & Griffin, J. A. Induction of RNA-stabilized DNA conformers by transcription of an immunoglobulin switch region. *Nature* **348**, 342–344 (1990).
20. Daniels, G. A. & Lieber, M. R. RNA:DNA complex formation upon transcription of immunoglobulin switch regions: implications for the mechanism and regulation of class switch recombination. *Nucleic Acids Res.* **23**, 5006–5011 (1995).
21. Tian, M. & Alt, F. W. Transcription-induced cleavage of immunoglobulin switch regions by nucleotide excision repair nucleases *in vitro*. *J. Biol. Chem.* **275**, 24163–24172 (2000).
22. Mizuta, R. *et al.* Molecular visualization of immunoglobulin switch region RNA/DNA complex by atomic force microscope. *J. Biol. Chem.* **278**, 4431–4434 (2003).
23. Tashiro, J., Kinoshita, K. & Honjo, T. Palindromic but not G-rich sequences are targets of class switch recombination. *Int. Immunol.* **13**, 495–505 (2001).
24. Dunnick, W., Hertz, G. Z., Scappino, L. & Gritzmacher, C. DNA sequences at immunoglobulin switch region recombination sites. *Nucleic Acids Res.* **21**, 365–372 (1993).
25. Shanmugam, A., Shi, M. J., Yauch, L., Stavnezer, J. & Kenter, A. L. Evidence for class-specific factors in immunoglobulin isotype switching. *J. Exp. Med.* **191**, 1365–1380 (2000).
26. Failli, A. *et al.* AID-dependent somatic hypermutation occurs as a DNA single-strand event in the BL2 cell line. *Nature Immunol.* **3**, 815–821 (2002).
27. Artsimovitch, I. & Landick, R. The transcriptional regulator RfaH stimulates RNA chain synthesis after recruitment to elongation complexes by the exposed nontemplate DNA strand. *Cell* **109**, 193–203 (2002).
28. Storb, U. *et al.* Cis-acting sequences that affect somatic hypermutation of Ig genes. *Immunol. Rev.* **162**, 153–160 (1998).
29. Yu, K., Chedin, F., Hsieh, C. L., Wilson, T. E. & Lieber, M. R. R-loops at immunoglobulin class switch regions within the chromosomes of stimulated B cells. *Nature Immunol.* (in the press).
30. Sartori, A. A., Schar, P., Fitz-Gibbon, S., Miller, J. H. & Jiricny, J. Biochemical characterization of uracil

processing activities in the hyperthermophilic archaeon *Pyrobaculum aerophilum*. *J. Biol. Chem.* **276**, 29979–29986 (2001).

31. Ramiro, A. R., Stavropoulos, P., Jankovic, M. & Nussenzweig, M. C. Transcription enhances AID-mediated cytidine deamination by exposing ssDNA on the nontemplate strand. *Nature Immunol.* (in the press).

Acknowledgements We thank T. Honjo for providing AID-deficient mice and B. Demple for helpful discussions. We also thank J. Manis, A. Zarrin, S. Ranganath and S. Saleque for suggestions and critical reading of the manuscript. This work was supported by a NIH grant (F.W.A.) and a NIH training grant (M.T.). K.C. is supported by a Pfizer Postdoctoral Fellowship in Rheumatology/Immunology and E.P. is supported by a Fondation pour la Recherche Medicale Postdoctoral Fellowship. F.W.A. is an Investigator and J.C. an Associate of the Howard Hughes Medical Institute.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to F.W.A. (e-mail: alt@enders.tch.harvard.edu).

Sequence-specific recruitment of transcriptional co-repressor Cabin1 by myocyte enhancer factor-2

Aidong Han*, Fan Pan†, James C. Stroud*, Hong-Duk Youn†, Jun O. Liu† & Lin Chen*

* Department of Chemistry and Biochemistry, University of Colorado at Boulder, Boulder, Colorado 80309-0215, USA

† Department of Pharmacology and Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA

The myocyte enhancer factor-2 (MEF2) family of transcription factors has important roles in the development and function of T cells, neuronal cells and muscle cells^{1–3}. MEF2 is capable of repressing or activating transcription by association with a variety of co-repressors or co-activators in a calcium-dependent manner^{1,4,5}. Transcriptional repression by MEF2 has attracted particular attention because of its potential role in hypertrophic responses of cardiomyocytes⁶. Several MEF2 co-repressors, such as Cabin1/Cain and class II histone deacetylases (HDACs), have been identified^{7–12}. However, the molecular mechanism of their recruitment to specific promoters by MEF2 remains largely unknown. Here we report a crystal structure of the MADS-box/MEF2S domain of human MEF2B bound to a motif of the transcriptional co-repressor Cabin1 and DNA at 2.2 Å resolution. The crystal structure reveals a stably folded MEF2S domain on the surface of the MADS box. Cabin1 adopts an amphipathic α -helix to bind a hydrophobic groove on the MEF2S domain, forming a triple-helical interaction. Our studies of the ternary Cabin1/MEF2/DNA complex show a general mechanism by which MEF2 recruits transcriptional co-repressor Cabin1 and class II HDACs to specific DNA sites.

Human MEF2B contains sequential MADS box (residues 2–58) and MEF2-specific (MEF2S) domain (residues 59–91), which are necessary and sufficient to bind a motif in human Cabin1 (residues 2156–2190)¹ (Fig. 1a). We determined the crystal structure of their complex bound to DNA with a MEF2 site by molecular replacement, using the previously determined MEF2A/DNA complex as a search model^{13,14}. The Cabin1 fragment and the carboxy-terminal part of the MEF2S domain that were absent from the starting model have a well-defined electron density (Fig. 1b). The overall structure of the Cabin1/MEF2B complex resembles a multi-layered pyramid built on the DNA (Fig. 1c, d). MEF2 forms a symmetric homodimer. Each MEF2 monomer consists of an extended amino-

terminal tail, three α -helices (H1, H2 and H3) and three β -strands (S1, S2 and S3). The N-terminal tail, helix H1 and strands S1 and S2 together form the MADS-box core domain, which mediates DNA recognition and dimerization in a manner similar to that observed in the MEF2A/DNA complexes^{14,15}. The MEF2S domain of each monomer forms a helix-strand-helix motif (H2-S3-H3) on the surface of the MADS-box core. At the top of the pyramid sits the α -helix of Cabin1 that forms a three-helix bundle with the H2 helices (Fig. 1c, d). Because the MADS box and part (helix H2) of the MEF2S domain observed in the MEF2A/DNA complex can be superimposed well on the corresponding part of the Cabin1/MEF2B/DNA complex (root-mean-square deviation 0.54 Å for 114 C α atoms; Fig. 2a), the binding of Cabin1 apparently has little effect on the structure and function of the MADS box of MEF2B.

A surprising structural feature of the ternary Cabin1/MEF2B/DNA complex is the entirely folded MEF2S domain, which was observed only partly (helix H2) in the MEF2A/DNA complexes with shorter MEF2 fragments^{14,15} (Fig. 2a). The observed structure of the MEF2S domain in this ternary complex is most probably due to intrinsic protein folding interactions rather than crystal packing and Cabin1 binding, because the same structure is found in the two

MEF2B monomers in different crystal packing environments. Moreover, Cabin1 makes no direct contact with H3 and interacts with S3 of each monomer differently, yet the structures of S3 of both monomers are identical. The MADS box in the transcription factor MCM1 and serum response factor (SRF) frequently binds its respective transcription partner through strand S2 and helix H1, which contain many exposed hydrophobic residues conserved between MCM1 and SRF^{16–18}. Remarkably, many of these residues, such as Met 29, Tyr 33, Phe 55 and Tyr 57, are also found in the MADS box of MEF2 (Fig. 1a). But in MEF2 the exposed hydrophobic surfaces of the MADS box interact with the MEF2S domain of the other monomer (Fig. 1c, d).

However, the detailed interactions between the MADS box and the MEF2S domain in MEF2B are strikingly similar to those used by the MADS box of MCM1 and SRF to interact with its respective partner^{16,17}. In the MEF2B dimer, strand S3 of one monomer forms parallel β -strands with S2 of the other monomer through main-chain hydrogen bonds and extensive side-chain interactions (Fig. 2b). Immediately after strand S3 is an amphipathic helix (H3) that docks perpendicularly on H1 of the opposing monomer through interactions that are primarily hydrophobic (Fig. 2c).

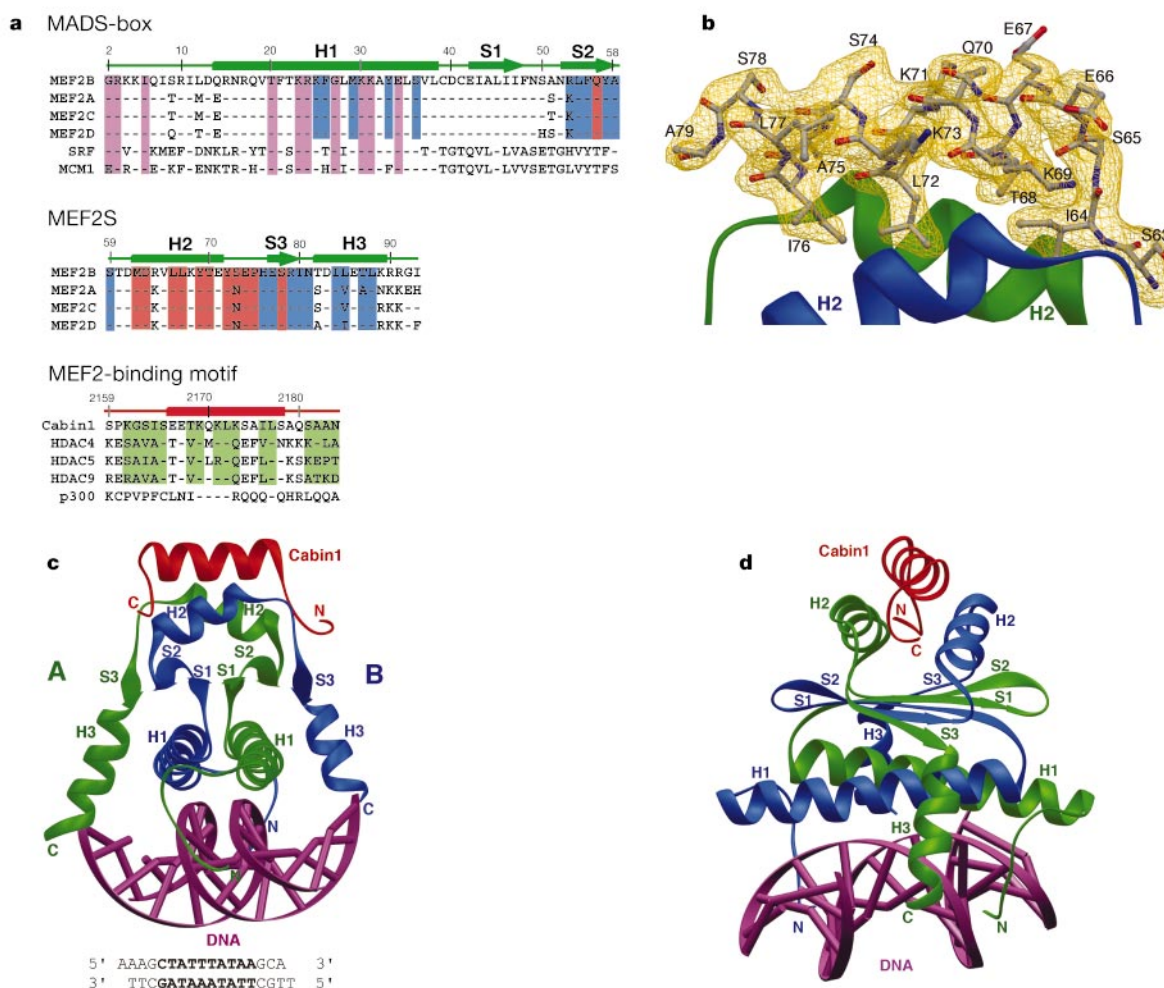


Figure 1 Structure of the Cabin1/MEF2B/DNA complex. **a**, Structure-based sequence alignments of the MADS box of MEF2A–D, SRF and MCM1 (top panel), the MEF2S domain of MEF2A–D (middle panel) and the MEF2-binding motif of Cabin1, HDAC4, HDAC5, HDAC9 and p300 (bottom panel). Residues in MEF2 involved in DNA binding (magenta), MADS-box–MEF2S interactions (blue) and Cabin1 binding (red) are coloured appropriately. MEF2-binding residues in Cabin1, HDAC4, HDAC5 and HDAC9 are

coloured green. α -Helices and β -strands are shown as bars and arrows, respectively. **b**, Simulated-annealing omit map of Cabin1. The 2.2-Å σ_A -weighted $F_0 - F_c$ map is contoured at 2.5 σ level. **c**, Overall structure of the Cabin1 (red), MEF2B (monomer A in green, monomer B in blue) and DNA (magenta) complex. **d**, The complex viewed at 90° to **c**.

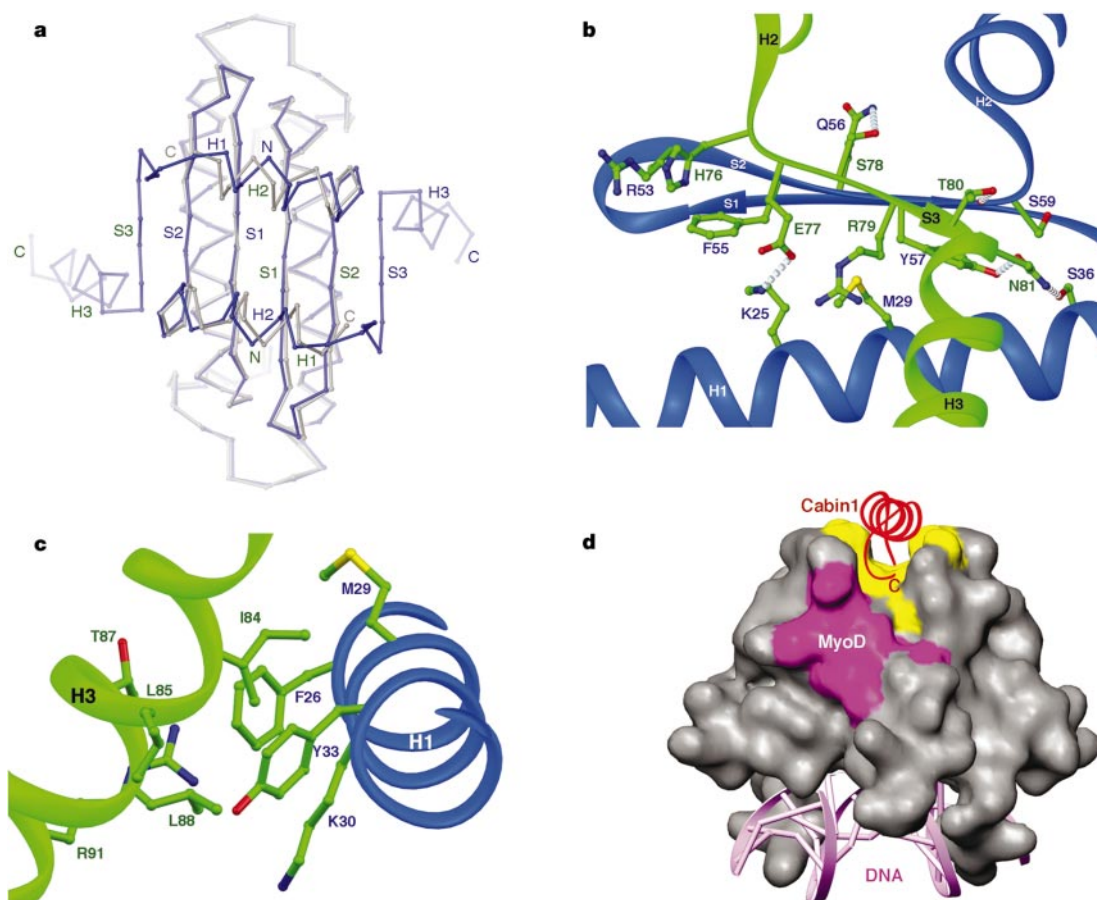


Figure 2 The fully folded structure of the MADS-box/MEF2S domain. **a**, Backbone superposition based on the C α atoms of residues 2–73 between MEF2A (both monomers are in grey) from the binary MEF2A/DNA complex¹⁴ and MEF2B (both monomers are in blue) from the Cabin1/MEF2B/DNA complex. **b**, Detailed interactions between β -strand S3 of monomer A with S2 of monomer B. The view is similar to that in Fig. 1d. In all figures,

residues from MEF2 are drawn in green and labelled with green and blue letters for monomers A and B, respectively. **c**, Detailed interactions between α -helix H3 of monomer A and H1 of monomer B. The view is similar to that in Fig. 1c. **d**, Surfaces mapped by mutagenesis as being important for interactions with MyoD (magenta) and class II HDACs (yellow)²⁰.

Residues mediating the interactions between the MEF2S domain and the MADS box are highly conserved in the MEF2 family (Fig. 1a), indicating that members of this family can form homodimers or possibly heterodimers with a similar structure.

The observed structure of the MEF2S domain has several functional implications. First, the MEF2S domain provides an interlock to stabilize dimerization further by interacting with the MADS box of the reciprocal subunit. As a result, a much larger surface area (6,228 Å²) is buried in the MEF2 dimer with an intact MEF2S domain (Fig. 1c, d). Second, the MEF2S domain could enhance DNA binding through direct stabilization of the DNA-binding helix H1, which is clamped down by helix H3 of the MEF2S domain¹⁹ (Figs 1c and 2c). Third, the MEF2S domain, which is unique to the MEF2 subfamily of the MADS-box proteins, might provide the major protein surface to interact with other transcription factors that 'crosstalk' with MEF2. Indeed, mutations on MEF2C (N73I/E74A/H76L or E77V/S78N/R79Q/T80A) have been shown to disrupt its ability to activate transcription synergistically with the MyoD family of transcription factors²⁰. These mutations map well to protein surfaces associated with strand S3 and its linker to S2 (Fig. 2d). Last, the fully folded MEF2S domain provides a stable docking site for the transcriptional co-repressor Cabin1 and potentially other transcriptional co-regulators.

Cabin1 is bound to the MEF2S domain in a deep groove with the central β -sheet as the floor and the two helices (H2) as the rim (Fig. 3a). The groove is lined with predominantly conserved

hydrophobic residues that engage in extensive van der Waals interactions with Cabin1 to form a binding site that is more extended than previously predicted^{14,15} (Fig. 3b). Additional β -strands (S3) of the MEF2S domain expand the central β -sheet of the MADS box to six strands (Figs 2a and 3a). A shorter MEF2B fragment (residues 2–78) lacking the C-terminal half of the MEF2S domain (S3 and H3) has much weaker affinity for Cabin1 than the construct used here (Supplementary Fig. 1). Overall, the Cabin1-binding site of MEF2 resembles the peptide-binding pocket of MHC, which also has a β -sheet floor and two helix rims²¹. However, unlike the extended peptide ligands bound by MHC, the MEF2-binding motif of Cabin1 adopts an α -helix to form a triple-helix bundle with MEF2.

The Cabin1 helix (residues 2166–2178) is amphipathic. Its hydrophobic face, including residues *Ile* 64, *Thr* 68, *Leu* 72, *Ile* 76 and *Leu* 77 (individual residues from Cabin1 are italicized throughout the text and are referred to by the last two digits of their actual sequence number), binds into the double-helix cleft of MEF2 (Fig. 4a). Located at the centre of the Cabin1/MEF2 interface and approximately on the dyad axis of the MEF2 dimer is *Leu* 72 of Cabin1, the side chain of which inserts deeply into a hydrophobic pocket formed by *Leu* 66, *Tyr* 69 and *Thr* 70 of each MEF2B monomer. N-terminal to *Leu* 72, *Ile* 64 and *Thr* 68 of Cabin1 together fill in a hydrophobic pocket formed by *Met* 62, *Leu* 66 and *Leu* 67 of monomer A and *Tyr* 69 of monomer B, whereas the dyad symmetry-related hydrophobic pocket is occupied by *Ile* 76

and Leu 77 of Cabin1. Thus, the MEF2-binding surface of Cabin1 has a pseudo-dyad symmetry that approximately matches the dyad symmetry of the MEF2 dimer, explaining the 1:2 binding ratio of Cabin1 to MEF2 observed in solution (Supplementary Fig. 2). The location of the hydrophobic pocket at the N terminus of helix H2 might account for the diagonal orientation of the Cabin1 helix (Figs 3a and 4a). The long aliphatic side chains of polar residues surrounding Leu 72, including Lys 69, Lys 71 and Lys 73, also make extensive van der Waals contacts with MEF2, further extending the binding interface (Figs 1b and 4b). The peptide segments of Cabin1 flanking its amphipathic helix bind MEF2 on both sides of the MADS-box/MEF2S domain by inserting into a surface groove formed by strands S2 and S3 and the linkers between S2, H2 and S3 (Supplementary Fig. 3). Overall, the Cabin1/MEF2B interface, burying 1,736 Å² solvent-accessible area, is largely hydrophobic and the intimate surface complementarities are probably the main determinants of the binding specificity between Cabin1 and MEF2 (Fig. 4b).

Our crystal structure is corroborated by mutagenesis. Mutations of key residues at the Cabin1/MEF2B interface, such as *Leu72Ala*, *Leu72Trp*, *Leu72Lys* or *Ile76Ala*, either completely abolished or significantly decreased the binding affinity of Cabin1 for MEF2B, whereas mutation of a Cabin1 residue outside the Cabin1/MEF2 interface, *Gln70Ala* or *Gln70Trp*, had little effect (Fig. 4c). Similar results were obtained when the interactions between the same set of Cabin1 mutants and MEF2B were determined by co-immunoprecipitation in whole-cell lysates of Jurkat T cells (Supplementary Fig. 5). It is remarkable that the removal of a three-carbon isopropyl group from the entire interface between Cabin1 and MEF2 in the

Leu72Ala mutant is sufficient to disrupt the Cabin1–MEF2 interaction, which is consistent with the key structural role of Leu 72 in the Cabin1/MEF2 interface (Fig. 4a and b).

The residues important for Cabin1 to bind MEF2B are conserved in a region of class II HDACs that are necessary and sufficient to bind MEF2 (refs 9, 10, 22, 23). The Cabin1-binding residues of MEF2B are also conserved in other members of the MEF2 family (Fig. 1a). A MEF2C triple mutant, V65A/L66S/L67R, failed to bind HDAC4 (ref. 24). In our crystal structure, Leu 66 and Leu 67 make extensive van der Waals contacts with *Ile 64*, *Thr 68*, *Ile 76* and *Leu 77* of Cabin1 (Fig. 4a), which are also conserved in class II HDACs (Fig. 1a). Moreover, *in vitro* binding data indicate that Cabin1 and HDAC4 bind MEF2 in a mutually exclusive manner (Supplementary Fig. 4). Taken together, the structural and biochemical data indicate that the triple-helix bundle observed in the Cabin1/MEF2B/DNA complex is a conserved structure adaptor for MEF2 to recruit transcriptional repressor Cabin1 or class II histone deacetylases to specific promoters.

Surprisingly, the V65A/L66S/L67R triple mutation on helix H2 has been found to diminish, rather than enhance, the transactivation function of MEF2 (ref. 19). A plausible explanation is that the

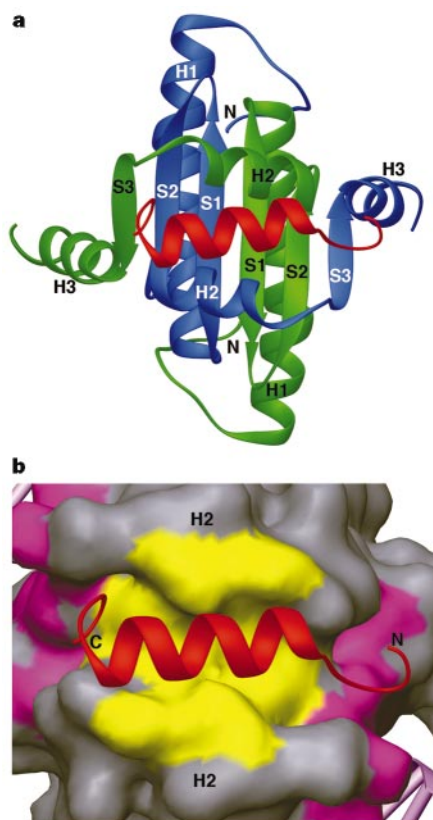


Figure 3 The overall structure of the Cabin1-binding site of MEF2B. **a**, The Cabin1-binding pocket composed of β -strands S1, S2, S3 and α -helix H2 of each monomer and the diagonal orientation of the Cabin1 helix in the pocket. **b**, A deep groove of the Cabin1-binding site lined with hydrophobic residues (the yellow patch).

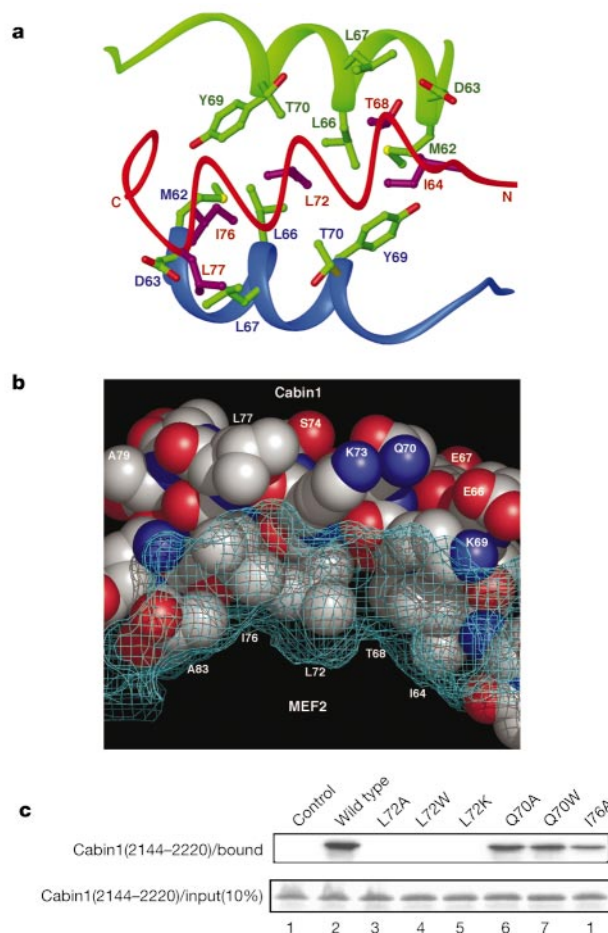


Figure 4 Crystallographic and mutational analyses of the Cabin1/MEF2B interface. **a**, Detailed interactions between the amphipathic helix of Cabin1 and the α -helix H2 of MEF2. Residues from Cabin1 (drawn in magenta) are labelled in red. **b**, The shape of Cabin1 (CPK model; C atoms in grey, N atoms in blue and O atoms in red) that complements the surface of MEF2 (mesh) well. **c**, *In vitro* binding of Cabin1 mutants to MEF2B. Top panel: wild-type Cabin1(2144–2200) (lane 2), *Leu72Ala* (lane 3), *Leu72Trp* (lane 4), *Leu72Lys* (lane 5), *Gln70Ala* (lane 6), *Gln70Trp* (lane 7) and *Ile76Ala* (lane 8). Bottom panel: input proteins used. Lane 1 is a control with beads only.

same helix in MEF2 is also involved in the recruitment of transcriptional co-activators. Indeed, the MADS-box/MEF2S domain has been shown to bind p300 competitively with Cabin1 and class II HDACs^{5,7}. A sequence search with the consensus MEF2-binding motif reveals a potential MEF2-binding site in a p300 fragment that binds MEF2 (ref. 25; Fig. 1a).

Our structural and biochemical studies define a new signalling domain conserved in the MEF2 family of transcription factors, whose peptide ligands might be present in a variety of signalling and transcription molecules capable of modulating MEF2-dependent transcription. There is great interest in the MEF2/class II HDAC pathway as a potential therapeutic target for heart hypertrophy⁶. Because DNA binding and co-regulator recruitment in MEF2 are mediated by protein surfaces located on opposite sides of the MADS-box/MEF2S domain, it is possible to inhibit the recruitment of transcriptional co-regulators by MEF2 without affecting the DNA binding of all MADS-box proteins¹⁹. This feature might be important for developing small-molecule ligands to modulate MEF2-dependent gene expression in clinical applications. □

Methods

Sample preparation and crystallization

Human MEF2B (residues 1–93) and Cabin1 (residues 2156–2190) were cloned in pET-30b as a fusion protein with Cabin1 at the C-terminus and a PreScission site in between. The protein was expressed in *Escherichia coli* BL21(DE3)pLysS and purified by ammonium sulphate precipitation and Sp-Sephadex chromatography. The purified fusion protein was cleaved with the PreScission Protease (Amersham Bioscience) and further purified by gel filtration. The DNA was prepared by solid-phase synthesis and purified by MonoQ under denaturing conditions. The DNA sequence is shown in Fig. 1c (MEF2-binding site in bold). The Cabin1/MEF2B/DNA ternary complex was prepared by mixing a 20% molar excess of DNA with the Cabin1/MEF2B complex and was further purified by Prep Cell (Bio-Rad Model 491) with an elution buffer of 5 mM Hepes, 30 mM NaCl, 0.5 mM EDTA, 1 mM dithiothreitol (DTT). To the complex peak was then added 20% excess DNA and it was concentrated to 0.4 mM for crystallization. Crystals were grown at 17 °C by the hanging-drop method with a reservoir buffer of 50 mM Bis-Tris propane (BTP), pH 6.35–6.68, 15% PEG 1000, 50 mM NaCl, 10 mM CaCl₂, 5 mM MgCl₂, 10% glycerol, 5 mM DTT and 2 mM spermine. Typically, cubic crystals grew to 300 µm in 2 weeks. The crystals belong to space group *P*4₁2₁2, with cell dimensions *a* = *b* = 70.14 Å, *c* = 151.88 Å.

Data collection, structure determination and analysis

Crystals were stabilized in harvest/cryoprotectant buffer (50 mM BTP, pH 6.68, 50 mM NaCl, 15% PEG 1000, 15% glycerol and 5 mM DTT). All crystals were flash-frozen in liquid nitrogen for storage and data collection under cryogenic conditions (100 K). The data were collected at the Advanced Photon Source (APS, Argonne National Laboratory) beam line (14-BM-C). Data were reduced with the programs DENZO and SCALEPACK²⁶. The structure of the ternary Cabin1/MEF2B/DNA complex was solved by the molecular replacement method by using the MEF2A/DNA binary complex as a partial search model¹⁴. Model building and refinement were performed with programs O and CNS^{27,28}. Throughout the refinement, non-crystallographic symmetry (NCS) restraints were applied to the MEF2 dimer. At the initial stage of the molecular replacement, the 'extra' electron density corresponding to the C-terminal half of the MEF2S domain and Cabin1 were clearly defined. The DNA in the crystals has two orientations; each was refined at half occupancy. The first two bases at the 5' end and the last base at the 3' end for both strands are disordered and are not included in the refined model. Water molecules were picked by using the automatic procedure implemented in CNS at the final stages of refinement²⁸. The final model has good geometry, as examined by PROCHECK²⁹. The *R*-factor of the refined model is 24.2%/31.3% (overall/last bin) and *R*_{free} is 26.8%/34.1% (overall/last bin). The model has 3,071 non-hydrogen atoms and 229 water molecules. Detailed statistics of the crystallographic analysis are presented in Supplementary Table 1. Diagrams of the structure were prepared with RIBBONS³⁰.

In vitro binding assay of Cabin1 mutants

The glutathione-S-transferase (GST)–MEF2B fusion was expressed in *E. coli* DH5α and purified as described previously¹. ³⁵S-labelled Cabin1 or its mutants was synthesized *in vitro* with pSG-Cabin1 (2144–2220) or the corresponding mutant plasmid (generated with the Stratagene site-mutagenesis kit) in a coupled transcription–translation system (TNT kit; Promega). Roughly equal amounts of wild-type and mutant proteins were incubated with a 50% slurry of GST–MEF2B immobilized on glutathione–Sepharose beads in 200–300 µl HEMG binding buffer (40 mM HEPES pH 7.8, 50 mM KCl, 5 mM MgCl₂, 0.1% Triton X-100, 10% glycerol, 1.5 mM DTT, protease inhibitors (Boehringer-Mannheim) and 0.5 mg ml^{−1} bovine serum albumin). The beads were washed four times with 1 ml (each time) HEMG buffer. Bound proteins were eluted in 30 µl of 50 mM Tris-HCl buffer, pH 7.4, containing 15 mM glutathione, resolved by SDS–polyacrylamide gel electrophoresis and detected by autoradiography. The control (Fig. 4c, lane 1) was performed with wild-type Cabin1 (2144–2220) and beads without immobilized GST–MEF2B.

Received 15 November 2002; accepted 7 March 2003; doi:10.1038/nature01555.

1. Youn, H. D., Sun, L., Prywes, R. & Liu, J. O. Apoptosis of T cells mediated by Ca²⁺-induced release of the transcription factor MEF2. *Science* **286**, 790–793 (1999).
2. Mao, Z., Bonni, A., Xia, F., Nadal-Vicens, M. & Greenberg, M. E. Neuronal activity-dependent cell survival mediated by transcription factor MEF2. *Science* **286**, 785–790 (1999).
3. Black, B. L. & Olson, E. N. Transcriptional control of muscle development by myocyte enhancer factor-2 (MEF2) proteins. *Annu. Rev. Cell Dev. Biol.* **14**, 167–196 (1998).
4. McKinsey, T. A., Zhang, C. L. & Olson, E. N. MEF2: a calcium-dependent regulator of cell division, differentiation and death. *Trends Biochem. Sci.* **27**, 40–47 (2002).
5. Youn, H. D., Grozinger, C. M. & Liu, J. O. Calcium regulates transcriptional repression of myocyte enhancer factor 2 by histone deacetylase 4. *J. Biol. Chem.* **275**, 22563–22567 (2000).
6. Zhang, C. L. *et al.* Class II histone deacetylases act as signal-responsive repressors of cardiac hypertrophy. *Cell* **110**, 479–488 (2002).
7. Youn, H. D. & Liu, J. O. Cabin1 represses MEF2-dependent Nur77 expression and T cell apoptosis by controlling association of histone deacetylases and acetylases with MEF2. *Immunity* **13**, 85–94 (2000).
8. McKinsey, T. A., Zhang, C. L., Lu, J. & Olson, E. N. Signal-dependent nuclear export of a histone deacetylase regulates muscle differentiation. *Nature* **408**, 106–111 (2000).
9. Sparrow, D. B. *et al.* MEF-2 function is modified by a novel co-repressor, MITR. *EMBO J.* **18**, 5085–5098 (1999).
10. Miska, E. A. *et al.* HDAC4 deacetylase associates with and represses the MEF2 transcription factor. *EMBO J.* **18**, 5099–5107 (1999).
11. Sun, L. *et al.* Cabin 1, a negative regulator for calcineurin signaling in T lymphocytes. *Immunity* **8**, 703–711 (1998).
12. Lai, M. M., Burnett, P. E., Wolosker, H., Blackshaw, S. & Snyder, S. H. Cain, a novel physiologic protein inhibitor of calcineurin. *J. Biol. Chem.* **273**, 18325–18331 (1998).
13. Woronicz, J. D. *et al.* Regulation of the Nur77 orphan steroid receptor in activation-induced apoptosis. *Mol. Cell. Biol.* **15**, 6364–6376 (1995).
14. Santelli, E. & Richmond, T. J. Crystal structure of MEF2A core bound to DNA at 1.5 Å resolution. *J. Mol. Biol.* **297**, 437–449 (2000).
15. Huang, K. *et al.* Solution structure of the MEF2A–DNA complex: structural basis for the modulation of DNA bending and specificity by MADS-box transcription factors. *EMBO J.* **19**, 2615–2628 (2000).
16. Tan, S. & Richmond, T. J. Crystal structure of the yeast MATα2/MCM1/DNA ternary complex. *Nature* **391**, 660–666 (1998).
17. Hassler, M. & Richmond, T. J. The B-box dominates SAP-1–SRF interactions in the structure of the ternary complex. *EMBO J.* **20**, 3018–3028 (2001).
18. Pellegrini, L., Tan, S. & Richmond, T. J. Structure of serum response factor core bound to DNA. *Nature* **376**, 490–498 (1995).
19. Molkenin, J. D., Black, B. L., Martin, J. F. & Olson, E. N. Mutational analysis of the DNA binding, dimerization, and transcriptional activation domains of MEF2C. *Mol. Cell. Biol.* **16**, 2627–2636 (1996).
20. Molkenin, J. D., Black, B. L., Martin, J. F. & Olson, E. N. Cooperative activation of muscle gene expression by MEF2 and myogenic bHLH proteins. *Cell* **83**, 1125–1136 (1995).
21. Bjorkman, P. J. *et al.* Structure of the human class I histocompatibility antigen, HLA-A2. *Nature* **329**, 506–512 (1987).
22. Lemerrier, C. *et al.* mHDA1/HDAC5 histone deacetylase interacts with and represses MEF2A transcriptional activity. *J. Biol. Chem.* **275**, 15594–15599 (2000).
23. Lu, J., McKinsey, T. A., Nicol, R. L. & Olson, E. N. Signal-dependent activation of the MEF2 transcription factor by dissociation from histone deacetylases. *Proc. Natl Acad. Sci. USA* **97**, 4070–4075 (2000).
24. Lu, J., McKinsey, T. A., Zhang, C. L. & Olson, E. N. Regulation of skeletal myogenesis by association of the MEF2 transcription factor with class II histone deacetylases. *Mol. Cell* **6**, 233–244 (2000).
25. Sartorelli, V., Huang, J., Hamamori, Y. & Kedes, L. Molecular mechanisms of myogenic coactivation by p300: direct interaction with the activation domain of MyoD and with the MADS box of MEF2C. *Mol. Cell. Biol.* **17**, 1010–1026 (1997).
26. Otwinowski, Z. in *Proceedings of the CCP4 Study Weekend* (eds Sawyer, L., Isaacs, N. & Burley, S.) 56–62 (SERC Daresbury Laboratory, Daresbury, UK, 1993).
27. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
28. Brunger, A. T. *et al.* Crystallography & NMR System: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54**, 905–921 (1998).
29. The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–776 (1994).
30. Carson, M. Ribbons 2.0. *J. Appl. Crystallogr.* **24**, 958–961 (1991).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements We thank H. Tong from APS beamline 14-BM, M. Giffin and D. Bates for help in data collection, G. Murphy, N. Ahn, J. Goodrich and D. Wuttke for critical reading of the manuscript, and T. McKinsey for discussion. This research was supported by a scholar award from the Damon Runyon–Walter Winchell Foundation (L.C.) and grants from the W. M. Keck foundation (L.C.) and the NIH (L.C. and J.O.L.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to L.C. (e-mail: lin.chen@colorado.edu). Coordinates have been deposited in the RCSB Protein Data Bank under accession code 1N6J.

Coordinated histone modifications mediated by a CtBP co-repressor complex

Yujia Shi*, Jun-ichi Sawada†‡, Guangchao Sui*†, El Bachir Affar*, Johnathan R. Whetstone*, Fei Lan*, Hidesato Ogawa‡, Margaret Po-Shan Luke*, Yoshihiro Nakatani‡ & Yang Shi*

* Department of Pathology, Harvard Medical School, 200 Longwood Avenue, Boston, Massachusetts 02115, USA

‡ Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA

† These authors contributed equally to this study

The transcriptional co-repressor CtBP (C-terminal binding protein) is implicated in tumorigenesis because it is targeted by the adenovirus E1A protein during oncogenic transformation¹. Genetic studies have also identified a crucial function for CtBP in animal development². CtBP is recruited to DNA by transcription factors that contain a PXL motif^{3,4}, but the detailed molecular events after the recruitment of CtBP to DNA and the mechanism of CtBP function in tumorigenesis are largely unknown. Here we report the identification of a CtBP complex that contains the essential components for both gene targeting and coordinated histone modifications, allowing for the effective repression of genes targeted by CtBP. Inhibiting the expression of CtBP and its associated histone-modifying activities by RNA-mediated interference resulted in alterations of histone modifications at the promoter of the tumour invasion suppressor gene E-cadherin and increased promoter activity in a reporter assay. These findings identify a molecular mechanism by which CtBP mediates transcriptional repression and provide insight into CtBP participation in oncogenesis.

To understand how CtBP represses transcription, we undertook the biochemical purification of CtBP⁵. We generated stable HeLa cells expressing human CtBP1 tagged with both the Flag and haemagglutinin (HA) epitopes at its amino terminus, which is expressed at a level comparable to that of the endogenous CtBP1 (Supplementary Fig. S2). Nuclear extracts from these cells were subjected to sequential purification with anti-Flag and anti-HA antibody columns. As shown in Fig. 1, about 20 polypeptides were found to be specifically associated with the tagged CtBP1 (compare lane 4 with lane 3), whose identities were determined by mass spectrometry (Fig. 1 and Supplementary Table 1). These CtBP1-associated proteins can be grouped into at least four classes based on their functions, which include the following: the DNA-binding proteins such as ZEB 1, which has previously been shown to interact with CtBP and to repress transcription in a CtBP-dependent manner⁶; the histone-modifying enzymes represented by the histone deacetylases HDAC1 and 2 and the related histone methyltransferases (HMTs) G9a and Eu-HMTase1 (EuHMT)^{5,7}; and two chromodomain-containing proteins HPC2 (ref. 8) and CDYL⁹. The last group includes CoREST^{10,11} (a protein found as a co-repressor to the transcription factor REST), a CoREST-related protein (KIAA1343) and a protein (KIAA0601) that shares sequence homology with polyamine oxidases, which we refer to as NPAO (nuclear polyamine oxidase). The presence of CoREST and its associated proteins in the CtBP complex suggests a possible interplay between these two co-repressors.

To define the CtBP complex further, we performed glycerol-gradient sedimentation and gel-filtration experiments. Most of the polypeptides listed above, namely CtBP2, G9a, EuHMT, CoREST, HDAC1 and 2, HPC2, NPAO, REBB-1, ZNF217 and KIAA0222, sedimented together with CtBP1 in fractions 17–19 (Fig. 2a), which was verified by western blot analysis (data not shown). Consistent

with this was the fact that gel-filtration chromatography of the purified CtBP complex followed by western blot analysis also identified the same set of proteins fractionating together with CtBP (Fig. 2b, peak fraction 22). Importantly, the native CtBP and its associated proteins also fractionated together in glycerol-gradient sedimentation assays (Supplementary Fig. S3). Taken together, these findings indicate that the above proteins are likely to be components of a core CtBP complex estimated to have a molecular mass of about $(1.3\text{--}1.5) \times 10^6$. To confirm these findings, we performed reciprocal immunoprecipitation. The G9a antibodies not only brought down G9a as expected but also immunoprecipitated EuHMT, CtBP1 and 2, HPC2, HDAC2 and CoREST together (Fig. 2c), showing that these proteins are indeed components of the same CtBP complex.

Both CtBP1 and CtBP2 were found in the core CtBP complex (Fig. 2), indicating that they might function as homodimers or heterodimers. In addition, significant amounts of CtBP1 and CtBP2 were also found in low-molecular-mass fractions preceding the CtBP complex (Fig. 2a and b). Although the exact nature of this pool of CtBP remains to be determined, they are likely to be CtBP multimers, given that CtBP1 and CtBP2 can form homodimers and heterodimers *in vitro*¹². CtBP might be present in two different pools in the cell—the free multimers and the complexed form—and the interchange between these two forms of CtBP might be a regulated event, as has been shown for CtBP interaction with certain DNA-binding transcription factors¹³.

Taken together, these results show the isolation of a CtBP complex that contains the essential elements for transcriptional repression, including components for promoter targeting and chromatin remodelling. The identification of both histone deacetylases and

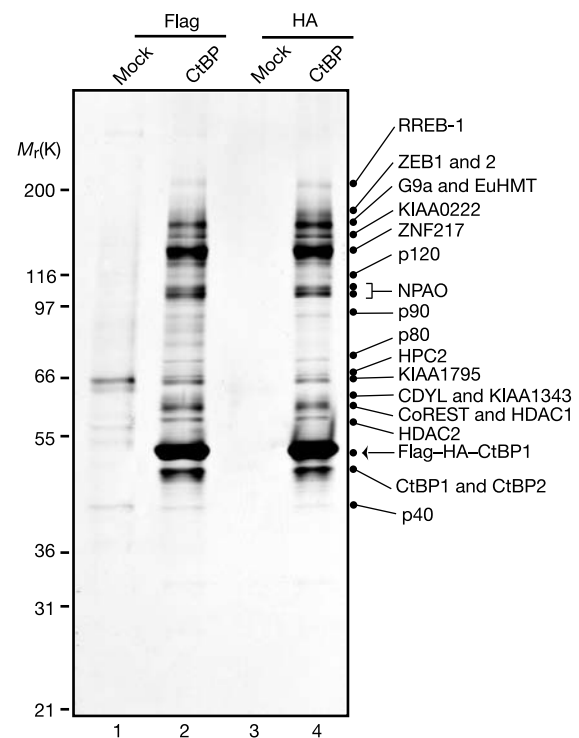


Figure 1 Purification and mass spectrometric analysis of polypeptides associated with CtBP1. Nuclear extracts from HeLa cells expressing human CtBP1 tagged with both Flag and HA epitopes were sequentially immunoprecipitated with the Flag and HA antibody affinity resins (lanes 2 and 4). Mock-transduced HeLa cells were used as a control (lanes 1 and 3). The CtBP-associated polypeptides were detected by silver staining. The identities of the associated polypeptides are indicated at the right.

methyltransferases suggests that the CtBP complex might function to coordinate stepwise histone modifications to result in a repressive chromatin environment.

To test the hypothesis that the CtBP complex coordinates histone modifications, we first analysed the CtBP complex for HDAC and HMT activities. The purified CtBP complex displayed significant and dose-dependent HDAC activity (Fig. 3a, lanes 5–7). We next analysed the CtBP complex for HMT activity, which has recently been shown to be involved in transcriptional repression^{14,15}. Core histones were incubated with either the CtBP complex or purified EuHMT⁵. Figure 3b shows that, as with EuHMT (lane 1), the purified CtBP complex can efficiently and specifically methylate histone H3 (lanes 2 and 3). The CtBP complex also showed a weak but significant HMT activity when mononucleosomes were used as substrates. These results show that the CtBP complex has HMT activity that can methylate both free and mononucleosomal histones *in vitro*.

We next wished to identify the sites on histone H3 that were methylated by the CtBP complex. The K9R (lysine → arginine) mutant was a significantly poorer methylation substrate than wild-type glutathione-S-transferase-linked histone H3 (Fig. 3c), indicating that Lys 9 is a major site for methylation. In contrast, mutation of Lys 4 had no effect on methylation. Although mutation of Lys 27 alone had no significant effect, double mutations of Lys 9 and Lys 27 completely eliminated methylation by the CtBP complex. These results suggest that, whereas Lys 9 is a major site, Lys 27 is a minor site for methylation by the CtBP complex *in vitro*.

Histone hyperacetylation is correlated with transcriptional activation, whereas histone hypoacetylation and methylation at Lys 9 of histone H3 are correlated with repression^{14,16,17}. Given that both HDAC and HMT activities are associated with CtBP, a model that might explain CtBP-mediated repression is that the CtBP complex can facilitate stepwise, coordinated, enzymatic reactions that

convert an active chromatin structure to a repressive one. We therefore asked whether the purified CtBP complex could directly convert Lys 9 acetylated histone H3 (active, AcK9-H3) to Lys 9 methylated H3 (repressive, MeK9-H3). We used histone H3 peptides pre-acetylated at either Lys 9 or Lys 14 as substrates and assayed them for methylation by the CtBP complex. As expected, AcK14-H3 peptides were efficiently methylated at Lys 9 (Fig. 4a, lanes 3 and 4). Significantly, AcK9-H3 peptides were also methylated by the CtBP complex (Fig. 4a, lanes 9 and 10). This methylation reaction was inhibited by sodium butyrate (Fig. 4a, compare lanes 9 and 10), whereas methylation of the unmodified histone H3 peptides by the CtBP complex was unaffected (Fig. 4a, compare lanes 3 and 4). In contrast, purified EuHMT, although methylating wild-type H3 peptide readily (Fig. 4a, lanes 1 and 2), failed to methylate AcK9-H3 (Fig. 4a, lanes 7 and 8). Taken together, our findings provide biochemical and functional evidence that a co-repressor complex, CtBP, possesses both HDACs and HMTs, which would allow it to convert chromatin directly from the active to the repressive state (Fig. 4b). These findings have thus uncovered a novel molecular mechanism by which CtBP represses transcription.

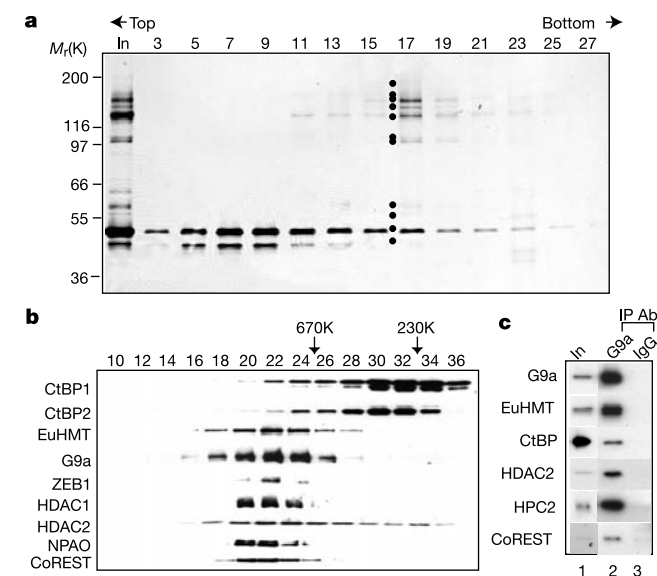


Figure 2 Characterization of a CtBP core complex. **a**, Glycerol gradient analysis. The CtBP complex material was separated on a 10–40% glycerol gradient by ultracentrifugation. Input (In) and fractions were detected by silver staining. Polypeptides that sedimented together within fraction 17 are highlighted by solid dots. **b**, Gel-filtration analysis. CtBP complex material was eluted from a Superose 6 HR gel-filtration column and analysed by western blotting. The numbers over the lanes represent the eluted fraction numbers, and protein molecular masses are indicated by arrows. **c**, Reciprocal immunoprecipitation (IP). Flag-purified CtBP complex was incubated with either anti-G9a (lane 2) antibody (Ab) or IgG (lane 3). The immunoprecipitates were subjected to western blot analysis. Lane 1, Flag-purified input material.

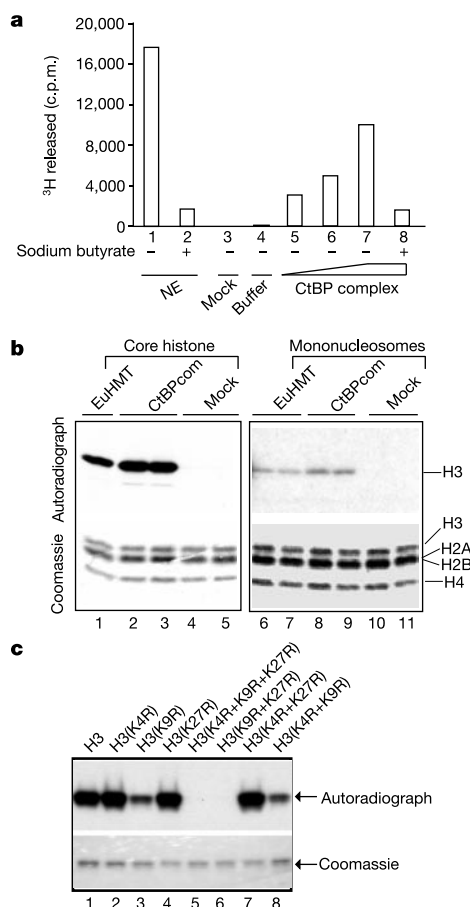


Figure 3 Enzymatic activities in the CtBP complex. **a**, HDAC activity. The ³H radioactivity released was measured in the presence (+) or absence (–) of sodium butyrate. NE, nuclear extract; mock, mock-purified material; buffer, reaction buffer alone. The amounts of CtBP complex used for reactions in lanes 5, 6, 7 and 8 were 1, 3, 10 and 10 μl, respectively. **b**, HMT activity. The HMT activity of the complex was measured with either core histone or mononucleosomes as substrate. EuHMT and mock-purified material were used as positive and negative controls. Lanes 2 and 3, 4 and 5, 6 and 7, 8 and 9, 10 and 11 are duplicates. **c**, Identification of methylation sites on histone H3 modified by the CtBP complex. Wild-type and a battery of glutathione-S-transferase-linked histone H3 mutants indicated on the top were used as substrates for methylation. The methylation sites were revealed by autoradiography.

To analyse further the functional significance of the association of HMTs with CtBP, we asked whether G9a and EuHMT are involved in CtBP-mediated repression. Previous studies suggested that ZEB1 and ZEB2 (SIP1) are involved in repression of the tumour invasion suppressor E-cadherin (E-cad)^{18,19}. The fact that the ZEB proteins are part of the CtBP complex makes it likely that CtBP is involved in repression of the E-cad promoter. If this is so, the removal of CtBP and its associated histone-modifying activities is expected to decrease the CtBP-mediated repression of E-cad transcription. We used the DNA vector-based RNA-mediated interference (RNAi) approach²⁰ to inhibit the expression of *ctbp1*, *g9a* and *eu-hmtase1* in E-cad-negative U2OS cells (Supplementary Fig. S4) as well as in E-cad-positive MCF-7 cells (data not shown). Inhibition of the expression of CtBP1, G9a and EuHMT resulted in a threefold to fourfold increase in the relative E-cad promoter reporter activity in U2OS cells (Fig. 4c), whereas the unrelated green fluorescent protein or LaminA/C short interfering RNAs²⁰ (siRNAs) had little effect on the reporter activity. Furthermore, the same set of siRNAs had little effect on the E-cad promoter reporter activity in MCF-7 cells (Fig. 4c). These data indicate that CtBP and the associated HMTs are involved in E-cad gene regulation.

To determine whether CtBP regulates histone modifications at the endogenous E-cad promoter, we performed chromatin immunoprecipitation (ChIP) experiments. We first investigated histone modifications at the E-cad promoter in U2OS and MCF-7 cells. Histone H3 was hypoacetylated and hypermethylated at Lys 9 in U2OS cells (Fig. 5a). In contrast, Lys 9 in MCF-7 cells was hyperacetylated but methylation was essentially undetectable. These findings are consistent with our hypothesis that CtBP represses the E-cad promoter, at least in part, by histone methylation at Lys 9 of histone H3. Our assays *in vitro* (Fig. 4a) indicate that CtBP-complex-mediated methylation of Lys 9 was preceded by histone deacetylation. Thus, blocking the action of histone deacetylases is predicted to result in a decrease in histone methylation and perhaps also an increase in acetylation at the CtBP target genes. Indeed,

treatment of cells with the HDAC inhibitor trichostatin A²¹ resulted in an increase in acetylation and a concomitant decrease in methylation at Lys 9 of histone H3 at the E-cad promoter (Fig. 5b).

We next used RNAi to determine whether CtBP is associated with methylation at Lys 9 of histone H3 at the E-cad promoter. CtBP1 expression in *ctbp1* siRNA-expressing cells was significantly decreased (Fig. 5c). We subsequently used these cells for ChIP and identified histone H3 Lys 9 methylation, CtBP1 and EuHMT at the E-cad promoter in U6 control HeLa cells (Fig. 5d, top panel, lanes 3–6). After inhibition of CtBP1 expression by RNAi, we observed a marked decrease in CtBP1 and EuHMT occupancy at the E-cad promoter (Fig. 5d, lower panel, lanes 4–6). These changes were accompanied by a significant decrease in methylation at Lys 9 of histone H3 (Fig. 5d, lower panel, lane 3). Taken together, these findings lend further support to the hypothesis that CtBP represses the E-cad promoter through the action of its associated histone-modifying activities.

On the basis of our findings, we propose that the action of the CtBP complex is initiated by the DNA-binding repressors that anchor the CtBP complex to its target promoters. This is followed by the action of HDAC1/2, which removes the acetyl group from the histone tails of transcriptionally active chromatin, allowing Lys 9 of histone H3 to be methylated by G9a and EuHMT. The methylated Lys 9 (and perhaps Lys 27) are then recognized by HPC2 and/or CDYL, which might contribute to the formation of a local repressive chromatin structure. These coordinated biochemical and enzymatic events help convert an active chromatin environment to one that is conducive to transcriptional repression. Biologically, the identification of the proteins involved in tumorigenesis in the CtBP complex^{18,19,22,23}, and the finding that CtBP and the associated HMTs repress transcription of the tumour invasion suppressor E-cad, shed significant light on the role of CtBP in tumorigenesis.

Last, we note that the CtBP complex might contain additional enzymes. We and others²⁴ have found that both CtBP complex and recombinant CtBP possess dehydrogenase activity (Supplementary Fig. S1). In addition, NPAO is related to polyamine oxidases, and CDYL shares homology with enoyl CoA isomerases/hydratases. We

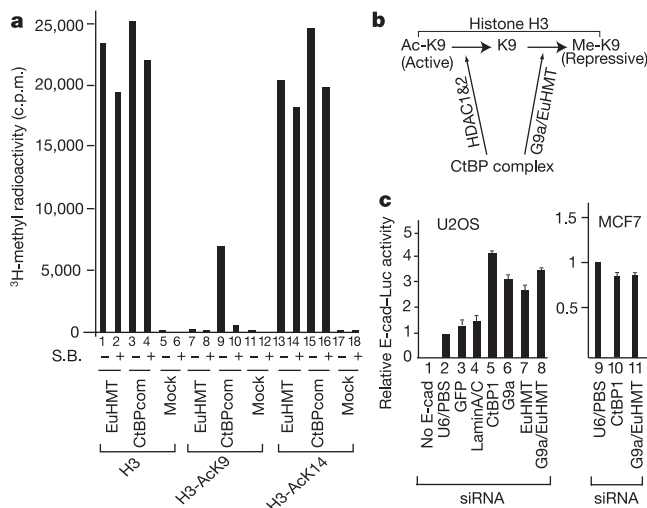


Figure 4 G9a and EuHMT are involved in CtBP-mediated repression. **a**, CtBP complex can directly convert AcK9-H3 to MeK9-H3. Various histone H3 peptides, including wild-type (H3), H3 with Lys 9 acetylated (AcK9-H3) or H3 with Lys 14 acetylated (AcK14-H3) were used as substrates for HMT assays of the CtBP complex (CtBPcom) in the absence (–) and presence (+) of sodium butyrate (S.B.). **b**, Schematic diagram of the CtBP complex activity. **c**, Knocking-down CtBP1, G9a or EuHMT expression by RNAi increases E-cad promoter activity. As indicated below each column, various siRNA constructs were co-transfected with the E-cad promoter (–289 to +5) luciferase reporter (E-cad–Luc) into U2OS cells or MCF-7 cells. The activity of the U6/PBS vector control was normalized as 1. Results are averages of three independent experiments.

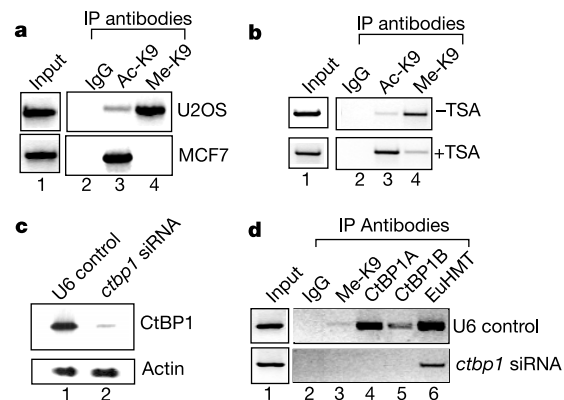


Figure 5 ChIP analysis of histone modifications at the endogenous E-cad promoter. **a**, H3-K9 methylation is correlated with a repressed E-cad promoter activity; ChIP analysis at the E-cad promoter in U2OS and MCF-7 cells using anti-AcK9-H3 (lane 3) and anti-MeK9-H3 (lane 4) antibodies. **b**, Inhibition of HDAC activity decreases K9-H3 methylation of E-cad promoter. U2OS cells treated with (+) or without (–) 100 ng ml^{–1} trichostatin A (TSA) were subjected to ChIP analysis for the E-cad promoter. **c**, *ctbp1* siRNA greatly decreased CtBP protein level. CtBP expression in the U6 control cells (lane 1) and in the *ctbp1*-siRNA cells (lane 2) was analysed by western blotting with an anti-CtBP1 mAb. **d**, *ctbp1* siRNA decreased CtBP complex occupancy and Lys 9 methylation at the E-cad promoter; ChIP analysis at the E-cad promoter in U6 control cells and *ctbp1* siRNA cells with the use of anti-CtBP1A (an affinity-purified rabbit polyclonal antibody), CtBP1B (SC-11390), EuHMT and MeK9-H3 antibodies. IP, immunoprecipitation.

speculate that, whereas HDACs and HMTs are involved in histone modifications, the other putative enzymes might also have a role in CtBP-regulated transcription, perhaps by targeting non-histone components of the chromatin. Future studies aimed at delineating their functions and possible relationships with histone-modifying enzymes in the CtBP complex are likely to provide new insight into eukaryotic gene regulation. □

Methods

CtBP1 complex purification

The detailed purification procedure has been described previously⁵. In brief, recombinant retroviruses expressing a bicistronic messenger RNA containing open reading frames of Flag–HA-tagged human CtBP1 and interleukin-2 receptor (IL-2R)- α was constructed and transduced into HeLa cells. The infected HeLa cells were sorted by anti-IL-2R monoclonal antibody (mAb) conjugated with magnetic beads, and the resulting Flag–HA–CtBP1 stable cell lines were propagated as suspension cells. Nuclear extract was made from 30 l of cells, from which the CtBP complex was purified by using anti-Flag M2 mAb-conjugated agarose beads (Sigma) followed by anti-HA 12CA5 mAb-conjugated agarose beads in buffer B (20 mM Tris-HCl, pH 7.9, 5 mM MgCl₂, 10% glycerol, 1 mM phenylmethylsulphonyl fluoride (PMSF), 0.1% Nonidet P40, 10 mM 2-mercaptoethanol) containing 100 mM KCl.

Mass spectrometry, glycerol gradient, gel filtration and reciprocal IP

Flag–HA double-purified material was separated by 4–20% gradient SDS–polyacrylamide-gel electrophoresis (SDS–PAGE) and stained with Coomassie blue. The various protein bands were excised and analysed by mass spectrometry at the Harvard Medical School Taplin Biological Mass Spectrometry Facility. For glycerol gradient sedimentation, 200 μ l of Flag–HA double-purified material or native HeLa nuclear extract was loaded on the top of a 5-ml 10–40% glycerol gradient cushion in buffer B containing 100 mM KCl and centrifuged for 10 h at 50,000 r.p.m (Beckman, SW55Ti). After centrifugation, individual fractions were collected from top to bottom, taking fractions of 100 μ l. Gel-filtration chromatography was performed with a Superose 6 HR 10/30 column (Amersham). CtBP complex material (200 μ l) was loaded on the equilibrated column and eluted with buffer B containing 100 mM KCl at flow rate of 0.25 ml min⁻¹, collecting fractions of 0.5 ml. For reciprocal immunoprecipitation, 50 μ l of Flag-purified CtBP-complex-containing material was incubated for 4 h with either affinity-purified anti-G9a antibody-conjugated or IgG-conjugated agarose beads at 4 °C. After extensive washing with buffer B containing 100 mM KCl, all associated proteins were eluted with 100 mM glycine-HCl buffer (pH 2.7) and subjected to western blotting.

HDAC and HMTase assays

The HDAC activity assay was performed using an HDAC assay kit (Upstate Biotechnology, Lake Placid, New York, USA). The HMTase activity assay was performed as described⁵. In brief, samples were incubated for 60 min at 30 °C in total of 20 μ l of methylation assay buffer (50 mM Tris-HCl, pH 8.5, 20 mM KCl, 10 mM MgCl₂, 10% glycerol, 10 mM 2-mercaptoethanol, 1 mM PMSF, 8 pmol S-adenosyl-[methyl-³H]methionine (specific radioactivity 60 Ci mmol⁻¹)) with either 50 μ g ml⁻¹ core histones, 100 μ g ml⁻¹ mononucleosomes or 20 μ g ml⁻¹ H3 peptide (residues 1–21) as substrate. Methylated histones or histone peptides were determined by filter-binding assays or radioactivity-excited fluorography after SDS–PAGE. For the HDAC-coupled HMTase activity assay, essentially the same assay condition as that described above was used, but the incubation time was increased to 2 h and the substrates were histone H3 peptides pre-acetylated at Lys 9 or Lys 14. All experiments were repeated at least three times.

Mammalian RNAi and cell-based transcriptional assays

The construction of siRNA-expressing vectors and indirect immunostaining were performed as described, as was the construction of control green fluorescent protein and LaminA/C siRNA plasmids²⁰. To establish a stable *ctbp1* siRNA cell line, we first subcloned a U6 promoter alone or U6-*ctbp1*-siRNA DNA fragment (from the above pBS *ctbp1* siRNA-expressing vector) into retroviral vector pMSCV – puro at the *Clal* site near the 3' long terminal repeat. These vectors were packaged into retroviruses, with which HeLa cells were then infected. Stable *ctbp1*-siRNA- or the U6 promoter-carrying cell lines were generated by selecting cells with 1.5 μ g ml⁻¹ puromycin. The ratio of RNAi to E-cad–Luc plasmids was adjusted to ensure that each cell that received the reporter DNA would also have received the RNAi plasmid. In a typical co-transfection experiment, 100 ng of the E-cad–Luc reporter DNA was co-transfected with 1.0 μ g of a siRNA-expressing plasmid and 50 ng of cytomegalovirus β -Gal plasmid into U2OS cells or MCF-7 cells in triplicate. Luciferase activity assays were performed 48–60 h after transfection and were normalized against β -galactosidase activity.

ChIP experiments

Chromatin preparation and ChIP experiments were performed in accordance with the

protocol from Upstate Biotechnology. Either a single-step or nested PCR reaction was used to amplify the E-cad promoter. The PCR conditions were optimized so that amplification was within the linear range for each primer pair. The following primers were used in this study: P1, 5'-TAGCCTGGCGTGGTGGTGTGCACCTG-3'; P2, 5'-GTGCGTGGCTGCAGCCAGGTGAGCC-3'; P3, 5'-CAGCTACTAGAGAGGCTGGGGCCAG-3'; P4, 5'-CGTACCCTGATTGGCTGAGGGTTC-3'; P5, 5'-GGGATTCGAACCCAGTGAA-3'; P6, 5'-CTGCGGCTCCAAGGGCCCA-3'. The amplicons were detected with ethidium bromide on a 2% agarose gel.

Received 21 January; accepted 4 March 2003; doi:10.1038/nature01550.

- Boyd, J. M. *et al.* A region in the C-terminus of adenovirus 2/5 E1a protein is required for association with a cellular phosphoprotein and important for the negative modulation of T24-ras mediated transformation, tumorigenesis and metastasis. *EMBO J.* **12**, 469–478 (1993).
- Chinnadurai, G. CtBP family proteins: More than transcriptional corepressors. *BioEssays* **25**, 9–12 (2003).
- Nibu, Y., Zhang, H. & Levine, M. Interaction of short-range repressors with *Drosophila* CtBP in the embryo. *Science* **280**, 101–104 (1998).
- Chinnadurai, G. CtBP, an unconventional transcriptional corepressor in development and oncogenesis. *Mol. Cell* **9**, 213–224 (2002).
- Ogawa, H., Ishiguro, K., Gaubatz, S., Livingston, D. M. & Nakatani, Y. A complex with chromatin modifiers that occupies E2F- and Myc-responsive genes in G0 cells. *Science* **296**, 1132–1136 (2002).
- Postigo, A. A. & Dean, D. C. ZEB represses transcription through interaction with the corepressor CtBP. *Proc. Natl Acad. Sci. USA* **96**, 6683–6688 (1999).
- Tachibana, M., Sugimoto, K., Fukushima, T. & Shinkai, Y. SET domain-containing protein, G9a, is a novel lysine-preferring mammalian histone methyltransferase with hyperactivity and specific selectivity to lysines 9 and 27 of histone H3. *J. Biol. Chem.* **276**, 25309–25317 (2001).
- Sewalt, R. G., Gunster, M. J., van der Vlag, J., Satijn, D. P. & Otte, A. P. C-terminal binding protein is a transcriptional repressor that interacts with a specific class of vertebrate Polycomb proteins. *Mol. Cell Biol.* **19**, 777–787 (1999).
- Lahn, B. T. & Page, D. C. Retroposition of autosomal mRNA yielded testis-specific gene family on human Y chromosome. *Nature Genet.* **21**, 429–433 (1999).
- You, A., Tong, J. K., Grozinger, C. M. & Schreiber, S. L. CoREST is an integral component of the CoREST-human histone deacetylase complex. *Proc. Natl Acad. Sci. USA* **98**, 1454–1458 (2001).
- Ballas, N. *et al.* Regulation of neuronal traits by a novel transcription complex. *Neuron* **31**, 353–365 (2001).
- Schmitz, F., Konigstorfer, A. & Südhof, T. C. RIBEYE, a component of synaptic ribbons: a protein's journey through evolution provides insight into synaptic ribbon function. *Neuron* **28**, 857–872 (2000).
- Zhang, Q. H., Piston, D. W. & Goodman, R. H. Regulation of corepressor function by nuclear NADH. *Science* **295**, 1895–1897 (2002).
- Zhang, Y. & Reinberg, D. Transcriptin regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Genes Dev.* **15**, 2343–2360 (2001).
- Kouzarides, T. Histone methylation in transcriptional control. *Curr. Opin. Genet. Dev.* **12**, 198–209 (2002).
- Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403**, 41–45 (2000).
- Jenuwein, T. & Allis, C. D. Translating the histone code. *Science* **293**, 1074–1080 (2001).
- Comijn, J. *et al.* The two-handed E box binding zinc finger protein SIP1 downregulates E-cadherin and induces invasion. *Mol. Cell* **7**, 1267–1278 (2001).
- Grootclaes, M. L. & Frisch, S. M. Evidence for a function of CtBP in epithelial gene regulation and anoikis. *Oncogene* **19**, 3823–3828 (2000).
- Sui, G. C. *et al.* A DNA vector-based RNAi technology to suppress gene expression in mammalian cells. *Proc. Natl Acad. Sci. USA* **99**, 5515–5520 (2002).
- Taunton, J., Hassig, C. A. & Schreiber, S. L. A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. *Science* **272**, 408–411 (1996).
- Collins, C. *et al.* Positional cloning of ZNF217 and NABCI: genes amplified at 20q13.2 and overexpressed in breast carcinoma. *Proc. Natl Acad. Sci. USA* **95**, 8703–8708 (1998).
- Satijn, D. P. *et al.* Interference with the expression of a novel human polycomb protein, hPc2, results in cellular transformation and apoptosis. *Mol. Cell Biol.* **17**, 6076–6806 (1997).
- Kumar, V. *et al.* Transcription corepressor CtBP is an NAD⁺-regulated dehydrogenase. *Mol. Cell* **10**, 857–869 (2002).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

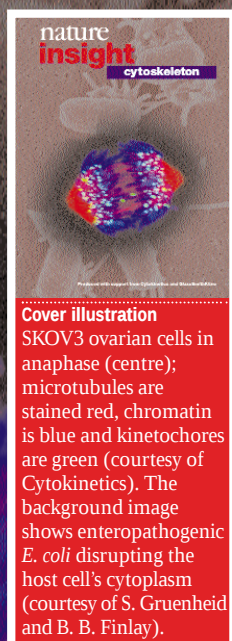
Acknowledgements We thank G. Gill, M. Greenberg and F. Winston for a critical reading of this manuscript, J. Parvin for advice, G. Chinnadurai for the CtBP1 cDNA, Y. Zhang for the glutathione-S-transferase–histone H3 fusion proteins, A. Otte and S. Tronick for antibodies, P. Silver for the parental retroviral vector, and D. Chen and S. Gygi for mass spectrometric analysis of the CtBP1 complex. Yu.S. was supported by a Program in Cancer Biology Training Grant from the National Cancer Institute. This work was supported by a grant from the NIH to Y.S.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to Y.S. (e-mail: yang_shi@hms.harvard.edu).

nature insight

cytoskeleton



The cytoskeleton of eukaryotic cells pervades the cytoplasm. It comprises three broad classes of proteins: actin filaments, microtubules and intermediate filaments. In addition to establishing cell and tissue shape, the cytoskeleton — along with associated motor proteins — influences a wide range of fundamental cellular functions, including cell migration, movement of organelles and cell division.

We are witnessing a rapid advance in our understanding of the cytoskeleton, driven in particular by determination of the structures of key molecules and acquisition of proteomics inventories of cytoskeletal proteins and their binding partners. The cytoskeleton is now no longer considered to be a rigid scaffold, but instead is viewed as a complex and dynamic network of protein filaments that can be modulated by internal and external cues.

This Insight examines many different facets of the cytoskeleton, reviewing the basic principles of filament organization, the operation of motor proteins and the role of the cytoskeleton in key biological processes. There is also consideration of the ways that pathogens subvert the cytoskeletal elements of the host cell to allow entry and spread of the invading organism. With this broad range of topics we aim to appeal not only to the cytoskeleton community, but also to the wide range of our readers who have an interest in cell biology.

Although significant progress has been made in understanding the cytoskeleton there is much still to be learnt. This *Nature* Insight, therefore, not only provides an overview of the current status of the field, but also provides perspectives on the directions of future research from leading scientists.

We are indebted to all the authors who contributed to the Insight and we apologize to those whose areas could not be covered owing to space restraints.

We are pleased to acknowledge the financial support of Cytokinetics and GlaxoSmithKline in producing this Insight. As always, *Nature* retains sole responsibility for editorial content and peer review.

Deepa Nath Senior Editor

overview:

- 741** The cytoskeleton, cellular motility and the reductionist agenda
T. D. Pollard

review articles:

- 746** Cell division
J. M. Scholey, I. Brust-Mascher & A. Mogilner
- 753** Dynamics and mechanics of the microtubule plus end
J. Howard & A. A. Hyman
- 759** Molecular motors
M. Schliwa & G. Woehlke
- 766** Adaptation of core mechanisms to generate cell polarity
W. J. Nelson
- 775** Microbial pathogenesis and cytoskeletal function
S. Gruenheid & B. B. Finlay

Editor, *Nature*: Philip Campbell
Insights Editors: Lesley Anson
Ursula Weiss
Consultant Editor: Christopher Surridge
Editorial Assistant: Simon Gibson

Art Director: Majo Xeridat
Layouts: Clifford Saunders
Diagrams: Ann Thomson
Suzanne Coleman
Debbie Maizels

Production Editor: Simon Gribbin
Production: Sue Gray
Marketing: Claire Aspinall
Sponsorship: Mark Greene

The cytoskeleton, cellular motility and the reductionist agenda

Thomas D. Pollard

Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven Connecticut 06520-8103, USA
(e-mail: thomas.pollard@yale.edu)

Eukaryotic cells depend on cytoskeletal polymers and molecular motors to establish their asymmetrical shapes, to transport intracellular constituents and to drive their motility. Cell biologists are using diverse experimental approaches to understand the molecular basis of cellular movements and to explain why defects in the component proteins cause disease. Much of the molecular machinery for motility evolved in early eukaryotes, so a limited set of general principles can explain the motility of most cells.

Three cytoskeletal polymers — actin filaments, microtubules and intermediate filaments (Table 1) — cooperate to maintain the physical integrity of eukaryotic cells and, together with molecular motors, allow cells to move themselves and their intracellular components. Although cellular motility has fascinated small groups of biologists for 300 years, interest in these processes has now spread to biologists more generally. The field has expanded as a result of insights gleaned about molecular mechanisms and the participation of cytoskeletal and motility molecules in many aspects of cellular function, including embryology, learning and memory, spread of cancer and microbial pathogenesis. The carefully regulated assembly of the cytoskeletal polymers and action of the associated motors is largely responsible for establishing cellular architecture and thus tissue structure.

This collection of reviews will bring readers up to date on several active areas of research. Howard and Hyman (page 753) explain how assembly and disassembly of microtubules produce forces to transport some intracellular molecules, chromosomes and organelles. Cellular locomotion powered by the assembly and disassembly of actin filaments¹ has many parallels with these microtubular mechanisms. Schliwa and Woehlke (page 759) cover the molecular motors that interact with actin filaments and microtubules to generate tension in the cytoskeleton as well as to move cargo as large as nuclei and as small as RNA molecules. Nelson (page 766) reviews how cells use cytoskeletal polymers and motors to generate asymmetry. Gruenheid and Finlay (page 775) cover the many ways that infectious organisms can hijack the motility system for their own purposes, while Scholey *et al.* (page 746) describe what we know about the segregation of chromosomes during mitosis and pinching daughter cells in two during cytokinesis.

These are spectacular examples of events where the cytoskeletal polymers and motors transiently assemble complex machines to carry out vital processes with high fidelity. The machines used for cellular locomotion, intracellular transport, mitosis and cytokinesis consist of millions of protein molecules held together by relatively weak, non-covalent bonds, which allows these machines to disassemble when their jobs are done, recycling their protein components for use at a later time.

In keeping with their fundamental contributions to cellular integrity and function, defects resulting from mutations in the genes for cytoskeletal and motility proteins cause human

disease. Recent examples include mutations in ankyrin (part of the membrane skeleton), which cause one type of cardiac arrhythmia², in titin in cardiomyopathies³, and in myosin-II in congenital defects of the brain and kidney⁴.

This perspective illustrates the power of the reductionist approach in cell biology and in studying the molecular basis of cellular movements in particular. Implementation of this agenda is based on three 'articles of faith'. First, owing to evolution from common ancestors, modern cells use a common set of molecular mechanisms to carry out their basic functions. Consequently, cell biologists believe that analysis of any experimentally tractable organism provides insights about general principles that will apply to most cells. Second, knowledge of the structures and functions of the individual parts of molecular machines reveals much about the workings of ensembles of molecules. And, third, a critical test for understanding is reconstitution of a complex process from purified components in 'wet' biochemical experiments and/or in computer simulations. Here I consider where the field stands with respect to these underlying beliefs and I conclude with a brief review of actin-based cellular motility, a topic not covered by the authors of the accompanying reviews.

Evolution

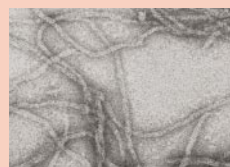
All five articles emphasize the contrast between the vast diversity of cellular behaviours and the unity of the underlying molecular mechanisms. Animal, plant and fungal cells differ remarkably in size, shape, motility and associations with other cells. Tiny yeast cells and most plant cells are trapped inside a cell wall, whereas animal cells can be either motile or confined to tissues by interactions with their neighbours. Most yeast segregate their chromosomes with a mitotic apparatus confined to the nucleus, whereas animals and plants have cytoplasmic mitotic apparatuses. Plants seem to rely largely on creation of a new plasma membrane and cell wall for cytokinesis, while both fungi and animals use a contractile ring of actin and myosin to divide.

Despite this diversity at the cellular level, the underlying mechanistic unity is now clear at the molecular level. All eukaryotic cells, in spite of their superficial differences, have inherited 'core mechanisms' (to quote Nelson) that are responsible for their structure and motility, including mitosis and cytokinesis. This core machinery appeared in a highly refined and effective form very early in the evolution of eukaryotes. Cells lacking this core machinery were lost. Given hundreds of millions of years since the main groups of

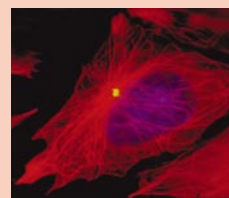
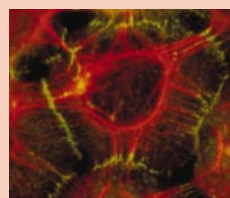
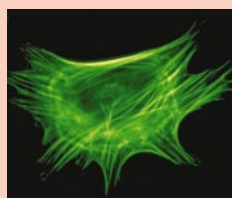
Table 1 Eukaryotic cytoskeletal polymers

Polymer	Actin filament	Microtubule	Intermediate filament
Protein subunit	Actin monomer	Tubulin heterodimer	Various proteins with an α -helical coiled-coil
Evolutionary origins	Prokaryotic hexokinase $\rightarrow$ prokaryotic actin-like proteins	Prokaryotic FtsZ	Early eukaryotic nuclear lamins
Polymerization by nucleation/elongation	Yes	Yes	Probably
Bound nucleotide	ATP	GTP	None
Ageing by nucleotide hydrolysis and phosphate release	Yes, allows binding of proteins that promote disassembly	Yes, destabilizes polymer	No
Flux of subunits through polymer at steady state (treadmilling)	Yes, very slow	Yes, slow	No
Dynamic instability (spontaneous fluctuations in length at steady state)	No	Yes, dramatic	No
Track for motors	Yes, 20 families of myosins	Yes, several dyneins and many families of kinesins	No

Electron micrographs of polymers



Fluorescence micrographs of cells with polymers



Micrographs reproduced with permission from ref. 31. The left fluorescence micrograph is from I. Herman, Tufts Medical School; the middle is from E. Smith and E. Fuchs, University of Chicago; and the right is from G. Borisy, University of Wisconsin.

eukaryotes separated from each other and given different selective pressures, their genomes have diverged significantly. A few genes for this core machinery were lost in specialized cells. Other genes acquired mutations that increased fitness for their organism's lifestyle. Some genes duplicated and then diverged to provide specialized functions. Although evolution refined the ancient mechanisms in each species, the core strategies are still used in contemporary cells that bear little superficial resemblance to each other. This allows investigators to search for general principles in those organisms that are most tractable for experimentation. Far from being an impediment, the diversity at the species level allows cell biologists to view the fundamental mechanisms of motility from a variety of perspectives.

Genes for actin and tubulin arose in prokaryotes⁵. Although the primary structures diverged extensively, crystal structures of prokaryotic actin-like and tubulin-like proteins are remarkably similar to their eukaryotic counterparts. Bacterial FtsZ binds GTP just like tubulin but polymerizes into long ribbons that participate in cytokinesis. Eukaryotic tubulin is a heterodimer of similar α - and β -subunits that assemble into cylindrical polymers (Table 1). The GTP bound to tubulin is hydrolysed and the γ -phosphate dissociates soon after incorporation of each tubulin molecule in a polymer. Dissociation of the γ -phosphate puts tubulin into a strained conformation that favours disassembly of the microtubules (see review by Howard and Hyman, page 753). Bacterial MreB binds ATP and forms actin-like filaments⁵ that are required for the elongated shape of rod-like bacteria. Some bacterial actins also help to partition DNA during mitosis⁶. (The assembly properties of actin are considered below.) In a fascinating role reversal early in eukaryotic evolution, actin filaments took over cytokinesis and microtubules assumed the partitioning of the genome.

Although actin filaments and microtubules differ in origin and structure, their shared features (Table 1) shows that evolution favoured extensive convergence of function. Moreover, nematodes

evolved completely different cytoskeletal polymers for their amoeboid sperm. Polymers of 'major sperm protein' lack any molecular similarity to actin, but carry out a cycle of assembly and disassembly that mimics that of actin in motile cells⁷.

Intermediate filaments arose during eukaryotic evolution rather than in prokaryotes and share little with the other cytoskeletal polymers. The rod-shaped protein subunits of intermediate filaments consist of a coiled-coil of α -helices and do not bind nucleotides. Owing to the symmetry of the subunits, the polymers are not polar like actin filaments and microtubules. Duplication and divergence of the genes for intermediate filament proteins produced a family of related genes in vertebrates. The protein products are expressed selectively in specialized cell types where they act as intracellular tendons that resist deformation of cells and tissues. Hair is composed of keratin intermediate filaments and illustrates the mechanical properties of these polymers. Mutations that interfere with the assembly of intermediate filaments result in mechanical fragility of the cells and tissues that depend upon them for their integrity. One example is mutations that compromise the keratin intermediate filaments in skin, cause blistering diseases⁸.

The molecular motors that move along microtubules and actin filaments had two origins. Dyneins are part of the family of AAA ATPases⁹ that also contribute to protein folding (Hsp100 chaperones), membrane traffic (*N*-ethylmaleimide-sensitive factor or NSF) and DNA synthesis (clamp loader proteins). The kinesin and myosin families of ATPase motors share a common core structure and may have the same common ancestor as the GTPases involved in signalling and protein synthesis¹⁰. Although GTPases are present in prokaryotes, compelling evidence for prokaryotic motors is still lacking.

The reductionist approach

Our understanding of the cytoskeleton and cellular motility is a triumph of the reductionist strategy, the approach that now dominates

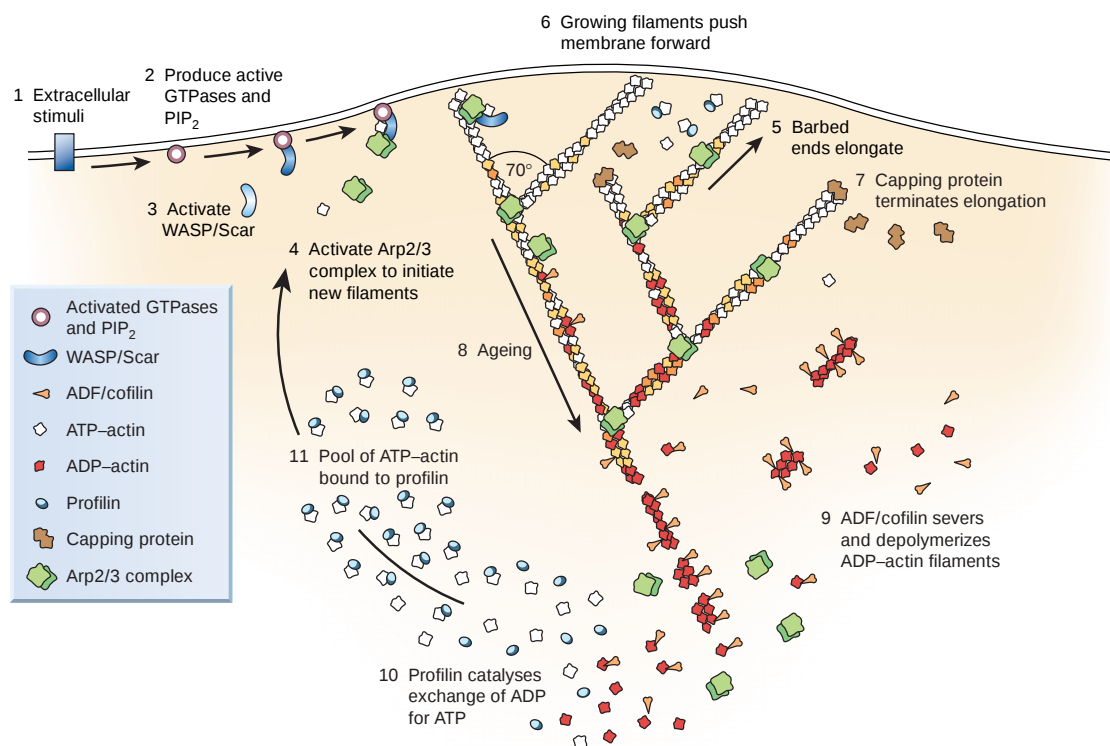


Figure 1 The dendritic-nucleation model for protrusion of lamellipodia. External cues (step 1) activate signalling pathways that lead to GTPases (2). These then activate Wiskott–Aldrich syndrome protein (WASP) and related proteins (3), which in turn activate Arp2/3 complex. Arp2/3 complex initiates a new filament as a branch on the side of an existing filament (4). Each new filament grows rapidly (5), fed by a high concentration of profilin-bound actin stored in the cytoplasm, and this pushes the plasma membrane forward (6). Capping protein binds to the growing ends, terminating elongation (7). Actin-depolymerizing factor (ADF)/cofilin then severs and depolymerizes the ADP filaments, mainly in the ‘older regions of the filaments’ (8, 9). Profilin re-enters the cycle at this point, promoting dissociation of ADP and binding of ATP to dissociated subunits (10). ATP-actin binds to profilin, refilling the pool of subunits available for assembly (11). (Image based on an original figure from ref. 32.)

research in cell biology. Sophisticated methods drive rapid progress, but we should aware of the limitations of these methods and the unfulfilled items on the reductionist agenda. The reductionist tasks include an inventory of the relevant molecules, determination of molecular structures, identification of molecular partners, measurement of rate and equilibrium constants for each reaction, localization of the molecules in live cells, physiological tests for participation in cellular processes and formulation of mathematical models to understand the system’s behaviour. Each review in this Insight section emphasizes parts of this agenda.

Reductionism starts with a list of the components. Most of the cytoskeletal proteins were discovered the ‘old-fashioned’ way, using purification by biochemical fractionation. Complete genome sequences and expressed sequence tag collections have expanded the inventory of cytoskeletal and motor proteins, particularly the diversity of isoforms of many of the proteins found in higher organisms. In a few cases experts have completed the annotation of selected genomes and defined the size of certain gene families such as myosins, which consists of more than 40 genes in humans¹¹. Similar work remains to be done for many other cytoskeletal gene families. Far less is known about the diversity of products generated by alternative splicing of pre-messenger RNAs.

Genetic screens and yeast two-hybrid assays have accelerated detection of protein partners, but traditional biochemical assays and affinity chromatography remain useful, particularly when empowered by sensitive analytical methods such as mass spectrometry. When scaled up to sample entire genomes or proteomes, these assays produce impressive interaction maps^{12,13}. Such efforts have saved an

immense amount of work and laid out a broad research agenda that is required to understand each interaction. These maps are, of course, a beginning rather than an end, as simple knowledge of an interaction will not explain how anything actually works.

Structure determines function, so the field eagerly awaits each new structure. Recent crystal structures include tubulin bound to a small regulator protein Op18/stathmin (see review in this issue by Howard and Hyman, page 753), bacterial actin and tubulin homologues⁵, and Arp2/3 complex (a seven-subunit nucleator of actin filaments¹⁴). Lacking crystals, three alternative approaches have yielded valuable structural information. First, Wiskott–Aldrich syndrome protein (WASP), a multi-domain protein that activates Arp2/3 complex, has been studied one domain at a time by nuclear magnetic resonance^{15,16}. Second, homology modelling based on other AAA ATPases was used to construct a preliminary model of dynein⁹. And third, technical advances in processing electron micrographs yielded an 8-Å structure of the microtubule¹⁷. Electron microscopy of single dynein molecules has recently led to a proposal for the mechanism of their ATP-driven power stroke¹⁸. Much work remains to complete a reference set of structures of cytoskeletal proteins.

Tracking the suspects

Light microscopy of live cells containing proteins tagged with fluorescent markers has revolutionized much of cell biology and replaced fluorescent antibody methods for many purposes. Expression of proteins fused to green fluorescent protein (GFP; and related proteins with different spectral properties) has made it possible to localize and study the dynamics of virtually any protein inside a living cell (and

even in tissues of live organisms; see review by Howard and Hyman, page 753, for examples). Investigators have embraced these methods with justifiable enthusiasm, but caution is required, as some fusion proteins cannot take the place of their wild-type counterparts in gene replacement experiments. Genetic manipulations make such controls routine in yeast laboratories, but they are rarely done in experiments on animal or plant cells.

Speckle microscopy has increased the power of fluorescent protein methods¹⁹. Expression of a low level of a GFP fusion protein or microinjection of a low concentration of purified protein labelled with a fluorescent dye leads to stochastic incorporation of labelled protein into microtubules, actin filaments or other cellular structures. The resulting speckles of fluorescence serve as fiduciary marks for orientation as the labelled structures move or turn over in live cells (see, for example, ref. 20).

Single-particle assays continue to make valuable contributions to understanding motility. One example is provided by the surprising solution to decades of controversy surrounding the mechanism of slow axonal transport. In this process, proteins such as the subunits of intermediate filaments move slowly (only 1–100 nm per second) from their site of synthesis in a neuronal cell body to the end of an axon or dendrite. Different experimental approaches gave apparently conflicting results regarding the movement of the molecules, whereas observation of single intermediate filaments revealed that they actually move rapidly but infrequently²¹. Propelled by motors, they move in fits and starts (but mostly stops) along microtubules.

Another example is bacteria that usurp the cytoplasmic actin system for propulsion through the cytoplasm of host cells. Observations of single bacteria and particles coated with bacterial proteins (or other activators) have defined the physics of the process²² and allowed reconstitution of the machinery from pure proteins²³. Similarly, much has been learned about the behaviour of microtubules²⁴ and actin filaments²⁵ by real-time observations of single polymers.

Knock downs and knock outs

Depletion of a protein from a cell remains the standard to assign function at the cellular level. Many laboratories continue these experiments one gene at a time using gene deletion in genetically tractable organisms. A complete set of deletion mutants for the budding yeast *Saccharomyces cerevisiae* has accelerated phenotyping. Depletion of mRNA and protein by RNA interference is faster, applicable to a growing range of cells and amenable to scaling up to screen the entire proteome for participation in a process such as cytokinesis (see review by Scholey *et al.*, page 746). However, in depletion experiments (as opposed to deletion experiments) one must keep in mind that severe reductions in concentration (or losses of affinity) may be required for physiological defects to appear; so false negatives are likely.

A complementary approach widely used in drug development and in a few academic laboratories is to screen target molecules or target cellular processes for inhibition with a library of small chemical compounds (for example, monasterol²⁶). ‘Chemical genetics’ or ‘chemical genomics’ are neologisms for the broadened scope of this traditional pharmacological approach. Given a library of sufficient size and diversity, it seems possible to find an inhibitor for most proteins. If specificity can be established, small-molecule inhibitors have exceptional value in analysing cellular processes, particularly if inhibition is reversible on a biologically relevant timescale of seconds to minutes.

Reaction mechanisms and systems properties

With some exceptions, the definition of reaction mechanisms still lags in most parts of this field. Chemical kinetics and measurements of force and motion of single molecules have established the mechanisms of several kinesins and myosins (see ref. 10, and review in this issue by Schliwa and Woehlke, page 759). This work is essential, because history has revealed repeatedly that mechanisms remain a

matter of speculation until cellular concentrations, affinities and reaction rates are known. Genetic interactions and identification of partners by semi- (or un-)quantitative precipitation assays are essential to initiate an investigation of mechanisms, but in every case known to me, the mechanism has turned out to be too complicated to understand without information about rates. Complete mechanisms are inevitably more interesting and pregnant with biological implications than superficial explanations.

Any cellular process involving more than a few types of molecules is too complicated to understand without a mathematical model to expose assumptions and to frame the reactions in a rigorous fashion. Second- and third-generation mathematical models are now being used to guide thinking and experimentation on the mechanisms of bacterial chemotaxis²⁷ and of the yeast cell cycle²⁸. The most advanced mathematical models in the field of cell motility deal with the actin filaments at the leading edge of continuously moving cells.

Cellular locomotion based on actin assembly

Primitive eukaryotes developed a mechanism to move towards food and away from harm that is based on the assembly of actin filaments (Fig. 1), which push the cell forward as the polymers grow at the leading edge of the cell (reviewed by ref. 1). All contemporary eukaryotes seem to use some variation of this ancient mechanism, although its manifestations vary from the movement of small ‘patches’ of actin filaments associated with the cell membranes of fungi to the rapid locomotion of cells such as human leukocytes. Genes required for this mechanism are found in protozoa, fungi, plants and animals. Although these genes are ancient, they have been conserved well enough through evolution that the protein parts seem to be fully interchangeable across species in biochemical assays.

Analysis of actin-based cellular motility illustrates how the reductionist strategy can be used to decipher a complex mechanism. So far, many of the key proteins have been identified and shown to reconstitute motility in a model system²³, all of their atomic structures are known, most of the rate and equilibrium constants have been measured, electron microscopy has revealed the organization of the machine in cells and a mathematical model correctly predicts the rate of movement²⁹.

Like tubulin, actin binds a nucleoside triphosphate, in this case ATP. After an actin molecule incorporates into a filament, the γ -phosphate is hydrolysed rapidly from the bound ATP. Dissociation of the γ -phosphate is slow, and ADP-actin has a lower affinity for the end of the filament, promoting dissociation and depolymerization.

This actin polymerization machine is intrinsically quiescent, but can be turned on by attractive chemical signals that direct cells such as protozoa, white blood cells or fibroblasts towards nutrients, prey or a tissue home. Acting through a variety of receptors, these cues activate signalling pathways that lead to small proteins that bind and hydrolyse GTP. These GTPases then activate proteins related to the product of the gene mutated in a human immunodeficiency disease called Wiskott–Aldrich syndrome. WASP and related proteins activate a large assembly of seven proteins called Arp2/3 complex, including two actin-related proteins (Arp2 and Arp3). Arp2/3 complex initiates a new filament as a branch on the side of an existing filament. Each new filament grows rapidly, fed by a high concentration of actin stored in the cytoplasm bound to the small protein profilin. Growth of the filaments pushes the plasma membrane (and the cell) forward. The energy comes from high-affinity binding of ATP-actin to the ends of filaments, similar to growing microtubules transporting cargo at their tips (discussed by Howard and Hyman, page 753). Initiation of new filaments as branches from the existing network provides a scaffold to push against.

The system is set up to terminate the growth of the filaments automatically before they grow so long that they do not push effectively and then to disassemble the network, so that the components can be recycled for an subsequent round of polymerization. First, capping protein binds to the growing ends, terminating elongation. Next, a

small protein called actin-depolymerizing factor (ADF)/cofilin binds weakly to the side of ADP-Pi actin filaments and promotes dissociation of the γ -phosphate. The ADP filaments become a target for higher-affinity binding of ADF/cofilin, leading to their severing and depolymerization. Profilin re-enters the cycle at this point, promoting dissociation of ADP and binding of ATP to dissociated subunits. ATP-actin binds to profilin, refilling the pool of subunits available for assembly.

Although many details of this mechanism remain unclear, a mathematical model incorporating both the molecular reactions and physical forces²⁹ correctly predicts the steady-state rate of cellular locomotion. This system has several advantages for modelling. It runs at steady state, the inventory of core proteins is small, the structures and concentrations of these proteins are known and biophysicists have measured many of the rate and equilibrium constants for the reactions. These models identify the variables that limit the rate of movement, such as the concentration of actin bound to profilin. In fact, when the concentration of unpolymerized actin is acutely lowered by releasing an actin-monomer sequestering protein locally in the cytoplasm, that part of a cell stops moving³⁰. The models raise a number of questions that can be addressed by further experimentation. Is the concentration of unpolymerized actin bound to profilin really the parameter limiting the rate of movement? Do interactions of the growing filaments with the inner surface of the membrane inhibit capping, thus biasing growth in the forward direction? How is the network of short, branched filaments remodelled into a network long unbranched filaments deeper in the cytoplasm?

Unmet challenges

Although we now have in hand a broad outline of the strategies that evolution has provided cells to produce motility and asymmetry, actual understanding of the physical mechanisms will require completion of the reductionist agenda. We still have gaps in our parts list and especially in biochemical mechanisms. As this agenda nears completion, the sheer complexity of most of the mechanisms driving cellular motility will force cell biologists to depend increasingly on mathematical models to test their hypotheses. Iterative cycles of quantitative modelling and quantitative experimentation are the only way to eliminate false but attractive hypotheses and to expose the valid features of models to rigorous scrutiny. Although rare in cell biology, this interplay of experiment and theory will gain in importance as the characterization of other systems advances. □

doi:10.1038/nature01598

1. Pollard, T. D. & Borisy, G. G. Cellular motility driven by assembly and disassembly of actin filaments. *Cell* **112**, 453–465 (2003).
2. Mohler, P. J. *et al.* Ankyrin-B mutation causes type 4 long-QT cardiac arrhythmia and sudden cardiac death. *Nature* **421**, 634–639 (2003).

3. Gerull, B. *et al.* Mutations of *TTN*, encoding the giant muscle filament titin, cause familial dilated cardiomyopathy. *Nature Genet.* **30**, 201–204 (2002).
4. Hu, A., Wang, F. & Sellers, J. R. Mutations in human nonmuscle myosin IIA found in patients with May-Hegglin anomaly and Fechtner syndrome result in impaired enzymatic function. *J. Biol. Chem.* **277**, 46512–46517 (2002).
5. van den Ent, F., Amos, L. A. & Lowe, J. Bacterial ancestry of actin tubulin. *Curr. Opin. Microbiol.* **6**, 634–638 (2001).
6. Møller-Jensen, J., Jensen, R. B., Löwe, J. & Gerdes, K. Prokaryotic DNA segregation by an actin-like filament. *EMBO J.* **21**, 3119–3127 (2002).
7. Roberts, T. M. & Stewart, M. Acting like actin. The dynamics of the nematode major sperm protein (msp) cytoskeleton indicate a push-pull mechanism for amoeboid cell motility. *J. Cell Biol.* **149**, 7–12 (2000).
8. Fuchs, E. & Cleveland, D. W. A structural scaffolding of intermediate filaments in health and disease. *Science* **279**, 514–519 (1998).
9. Mocz, G. & Gibbons, I. R. Model for the motor component of dynein heavy chain based on homology to the AAA family of oligomeric ATPases. *Structure* **9**, 93–103 (2001).
10. Vale, R. D. & Milligan, R. A. The way things move: looking under the hood of molecular motors proteins. *Science* **288**, 88–95 (2000).
11. Berg, J. S., Powell, B. C. & Cheney, R. E. A millennial myosin census. *Mol. Biol. Cell* **780**–794 (2001).
12. Tong, A. H. *et al.* A combined experimental and computational strategy to define protein interaction networks for peptide recognition modules. *Science* **295**, 321–324 (2002).
13. Tong, A. H. *et al.* Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* **294**, 2364–2368 (2001).
14. Robinson, R. C. *et al.* Crystal structure of Arp2/3 complex. *Science* **294**, 1660–1661 (2001).
15. Kim, A. S., Kakalis, L. T., Abdul-Manan, N., Liu, G. A. & Rosen, M. K. Autoinhibition and activation mechanisms of the Wiskott–Aldrich syndrome protein. *Nature* **404**, 151–158 (2000).
16. Volkman, B. F., Prehoda, K. E., Scott, J. A., Peterson, F. C. & Lim, W. A. Structure of the N-WASP EVH1 domain-WIP complex: insight into the molecular basis of Wiskott–Syndrome. *Cell* **111**, 565–576 (2002).
17. Li, H., DeRosier, D. J., Nicholson, W. V., Nogales, E. & Downing, K. H. Microtubule structure at 8 Å resolution. *Structure* **10**, 1317–1328 (2002).
18. Burgess, S. A., Walker, M. L., Sakakibara, H., Knight, P. J. & Oiwa, K. Dynein structure and power stroke. *Nature* **421**, 715–718 (2003).
19. Waterman-Storer, C. M., Desai, A., Bulinski, J. C. & Salmon, E. D. Fluorescent speckle microscopy, a method to visualize the dynamics of protein assemblies in living cells. *Curr. Biol.* **8**, 1227–1230 (1998).
20. Watanabe, N. & Mitchison, T. J. Single-molecule speckle analysis of actin filament turnover in lamellipodia. *Science* **295**, 1083–1086 (2002).
21. Wang, L. & Brown, A. Rapid intermittent movement of axonal neurofilaments observed by fluorescence photobleaching. *Mol. Biol. Cell* **12**, 3257–3267 (2001).
22. Gerbal, F., Chaikin, P., Rabin, Y. & Prost, J. An elastic analysis of *Listeria monocytogenes* propulsion. *Biophys. J.* **79**, 2259–2275 (2000).
23. Loisel, T. P., Boujemaa, R., Pantaloni, D. & Carlier, M. F. Reconstitution of actin-based motility of *Listeria* and *Shigella* using pure proteins. *Nature* **401**, 613–616 (1999).
24. Walker, R. A. *et al.* Dynamic instability of individual microtubules analyzed by video light microscopy: rate constants and transition frequencies. *J. Cell Biol.* **107**, 1437–1448 (1988).
25. Amann, K. J. & Pollard, T. D. Direct real-time observation of actin filament branching mediated by Arp2/3 complex using total internal reflection microscopy. *Proc. Natl Acad. Sci. USA* **98**, 15009–15013 (2001).
26. Peterson, J. R. & Mitchison, T. J. Small molecules, big impact. A history of chemical inhibitors and the cytoskeleton. *Chem. Biol.* **9**, 1275–1285 (2002).
27. Bray, D. Bacterial chemotaxis and the question of gain. *Proc. Natl Acad. Sci. USA* **99**, 7–9 (2002).
28. Tyson, J. J., Chen, K. & Novak, B. Network dynamics and cell physiology. *Nature Rev. Mol. Cell Biol.* **2**, 908–916 (2001).
29. Mogilner, A. & Edelstein-Keshet, L. Regulation of actin dynamics in rapidly moving cells: a quantitative analysis. *Biophys. J.* **83**, 1237–1258 (2002).
30. Roy, P. *et al.* Local photorelease of caged thymosin β 4 in locomoting keratocytes causes cell turning. *J. Cell Biol.* **153**, 1035–1048 (2002).
31. Pollard, T. D. & Earnshaw, W. C. *Cell Biology* (W. B. Saunders, New York, 2002).
32. Pollard, T. D., Blanchoin, L. & Mullins, R. D. Biophysics of actin filament dynamics in nonmuscle cells. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 545–576 (2000).

Cell division

Jonathan M. Scholey, Ingrid Brust-Mascher & Alex Mogilner

Laboratory of Cell and Computational Biology, Center for Genetics and Development, University of California, Davis, California 95616, USA
(e-mail: jmscholey@ucdavis.edu)

In creating the mitotic spindle and the contractile ring, natural selection has engineered fascinating precision machines whose movements depend upon forces generated by ensembles of cytoskeletal proteins. These machines segregate chromosomes and divide the cell with high fidelity. Current research on the mechanisms and regulation of spindle morphogenesis, chromosome motility and cytokinesis emphasizes how ensembles of dynamic cytoskeletal polymers and multiple motors cooperate to generate the forces that guide the cell through mitosis and cytokinesis.

During the nineteenth century, the discovery that cells reproduce themselves by dividing into two illuminated the very origin of cells and became a cornerstone of the cell theory^{1,2}. Today, research on cell division flourishes because an improved understanding of its mechanism could lead to improvements in the treatment of diseases such as cancer³ and because we are fascinated by the cytoskeletal 'nanomachinery' that is responsible for mitosis and cytokinesis^{4–14}.

The pathways by which the microtubule (MT)-based mitotic spindle and the actin-based contractile ring use cytoskeletal proteins to coordinate mitosis and cytokinesis are well understood^{4–14} (Fig. 1). During mitosis, the spindle uses MTs and multiple mitotic motors to distribute identical copies of the replicated genome to the products of each division^{2,4–9}. Usually this process begins during prophase (Fig. 1a) with the migration of duplicated centrosomes around the nuclear envelope. The envelope breaks down at the onset of prometaphase, allowing spindle MTs to capture chromosomes and move them to the equator (congression; Fig. 1b), so that by metaphase (Fig. 1c), pairs of sister chromatids lie on the spindle equator facing opposite spindle poles. Upon the onset of anaphase¹⁰, cohesion between sister chromatids is lost, which allows sister chromatids to be moved to opposite spindle poles (anaphase A; Fig. 1d) as the spindle poles themselves move further apart (anaphase B; Fig. 1e). Also during anaphase, the spindle delivers a signal to the cortex (Fig. 1d inset) that defines the position and orientation of the contractile ring, the machine that uses actin and myosin-II to drive cytokinesis (Fig. 1e inset)¹¹. The contraction of this ring causes the furrow to ingress as the nuclear envelopes reassemble around sets of decondensing segregated sisters. Finally, the furrow 'seals', completing the separation of the daughter cells (Fig. 1f).

Cells use a significant fraction of their proteins to divide — functional proteomics¹² indicates that *Caenorhabditis elegans* uses 6% of its open reading frames to encode proteins required for cell division and an important subset of these proteins comprise actin filaments, MTs, motor proteins and accessory proteins^{13,14}. MTs and actin filaments are linear, polar, multistranded polymers, built from 13 strands of $\alpha\beta$ -tubulin heterodimers and 2 strands of G-actin monomers, respectively. These polymers can generate pushing and pulling forces as they grow and shrink by addition and loss of subunits from their ends, and they also serve as tracks for motor proteins that use ATP hydrolysis to generate force and motility⁷ (Box 1). At the single-molecule level, cytoskeletal proteins generate piconewton-scale forces and nanometre-

scale movements^{7,13,14}, but during cell division they function as ensembles that are capable of generating forces in the range of nanonewtons and serve to accurately move intracellular components and rearrange areas of the cell surface over distances of tens of microns^{7,13–17}. How do these cytoskeletal force generators cooperate to drive the motility events underlying the mechanics and regulation of cell division?

Spindle morphogenesis and elongation

The purpose of mitosis is to segregate sister chromatids by moving them to opposite poles. To this end, spindle MTs become oriented into a bipolar array whose dyad axis divides the structure into two half spindles (Fig. 1c). Within each half-spindle, the MTs lie on trajectories that point their minus-ends towards a focus at the poles, allowing spindle forces to accomplish their goal by translocating chromatids along these trajectories (Fig. 1d). Bipolar spindles can form by two pathways, the centrosome-directed assembly pathway⁴, in which MT assembly is nucleated by centrosomes, or the chromosome-directed pathway^{5,6}, in which chromosomes induce MT assembly (Fig. 2a). The relationship between these pathways is unresolved, as is the question of why some cells (such as *Drosophila* embryos) use the centrosome-directed pathway whereas others (for example, *Drosophila* female oocytes) lack centrosomes and use the alternate chromosome-directed pathway.

Centrosomes consist of a pair of cylindrical centrioles surrounded by pericentriolar material that contains the MT-nucleating γ -tubulin ring complex (γ -TuRC). Electron microscopy suggests that the γ -TuRC acts as a helical template for new MTs which grow by subunit addition at their plus ends^{18–21}. Recent work suggests that a MT-associated protein (MAP) called XMAP215 is important for MT nucleation at centrosomes²². Perhaps the γ -TuRC and XMAP215 play complementary roles, stabilizing lateral and longitudinal bonds between subunits, respectively²².

In the chromosomal pathway, the guanine nucleotide-exchange factor of the small GTPase, Ran, generates a spatial gradient of active Ran-GTP around chromosomes²³. Ran-GTP promotes the release of factors that induce MT assembly from a pool that is sequestered in an inactive form by importin- β , thereby activating spindle assembly around chromosomes²⁴. Spindles that lack conventional centrosomes may contain 'pseudo-centrosomes' consisting of various MAPs that are important for spindle pole formation and stability²⁵. One of these MAPs is XMAP215, which may be transported to the poles by a minus-end-directed C-terminal kinesin, where it could nucleate MT assembly as in the centrosome-directed pathway²⁵.

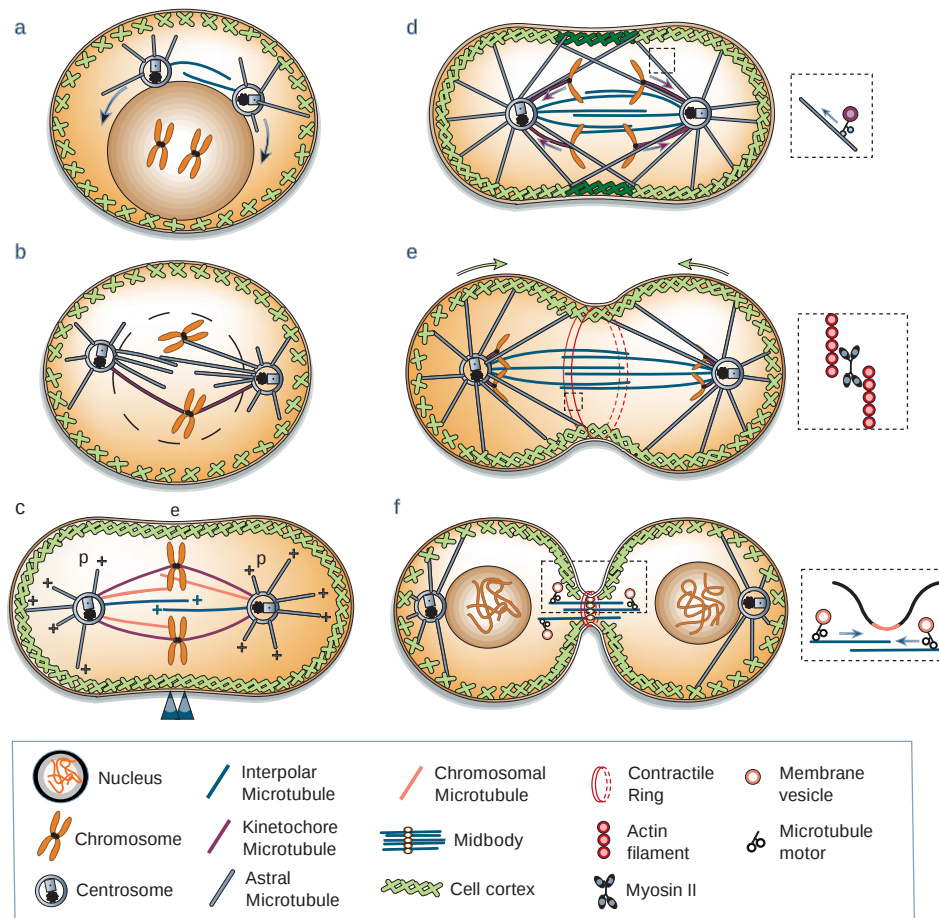


Figure 1 Mitosis and cytokinesis. **a**, Prophase. Duplicated centrosomes migrate around the nucleus. (Centrosomes, consisting of a pair of previously replicated centrioles surrounded by pericentriolar material, nucleate MT assembly and organize spindle poles.) **b**, Prometaphase. The nuclear envelope breaks down allowing MTs to move chromosomes to the equator (e) in a process termed congression. **c**, Metaphase. Sister chromatids (double arrowheads) face opposite poles (p). MTs are oriented with their plus ends distal to the poles, and are organized into four sets, namely: astral MTs, which link spindle poles to the cell cortex; chromosomal MTs, which link chromosome arms to poles; kinetochore MTs (kMTs), which link poles to kinetochores; and interpolar MTs (ipMTs), which link the two poles. **d**, Anaphase A. Chromatids are moved to opposite poles (segregation). **e**, Anaphase B. Pole–pole spacing increases. During late anaphase the division plane is determined by a mechanism involving spindle–cortex interactions and the cleavage furrow containing a contractile ring assembles from actomyosin-II and begins to contract. **f**, Telophase/cell–cell scission. Nuclear envelopes reassemble around decondensing segregated sisters. The contractile ring contracts (furrow ingression) developing a barrier between the daughter cells and constricting the spindle mid-zone (the array of ipMTs lying between separated chromatids) into a structure called the midbody (the remnant of the mid-zone). During abscission, the furrow ‘seals’ by a mechanism thought to involve vesicle transport/exocytosis, completely separating the daughter cells.

MT motors have subtly different roles in the centrosome- and chromosome-directed assembly pathways^{4–6}. In the latter case, MTs randomly organized around chromosomes become crosslinked into antiparallel bundles by bipolar kinesins, then plus-end-directed chromokinesins reorganize these MTs to position their minus ends distal to chromosomes, and finally minus-end-directed motors (dynein or Ncd) crosslink the MT minus ends into focused poles²⁶. In the former case, duplicated centrosomes are moved apart by shifts in a balance of outward and inward forces generated by the cooperative action of dynamic MTs, cortical dynein and multiple MT sliding motors localized to interpolar MT (ipMT) bundles^{4,27}. Bipolar (plus-end-directed) and C-terminal (minus-end-directed) kinesins acting on ipMTs are candidates for generating some of the antagonistic outward and inward forces that position spindle poles⁴. Recent computer simulations have suggested that bipolar and C-terminal kinesins could generate outward and inward forces on the poles, as expected, but various mixtures of these motors were unable to produce a robust isometric spacing of two spindle poles unless the two kinesins were organized into co-polymers²⁸.

The formation of such motor co-polymers might be one function of the hypothetical ‘spindle matrix’ that is proposed to serve as a substrate for the organization and activity of MTs and motors^{29,30}. Another possible function of this matrix, if it exists, is to strengthen the spindle machinery, because physical estimates suggest that the large forces developed by spindles (in the nanonewton range)¹⁵ would cause MT buckling unless the MTs are stabilized by a matrix and/or by MT–MT crosslinkers⁷. Definitive evidence for a spindle matrix is lacking, but perhaps matrix proteins are lurking undetected in spindle pole and mitotic MT preparations, which can be characterized by powerful matrix-assisted laser desorption/ionization mass spectroscopy^{31,32}.

Spindle MTs are highly dynamic (Fig. 2b) and several accessory proteins influence spindle MT dynamics (Fig. 2c). For example, the MAP XMAP215, which promotes MT polymerization, and the kinesin XKCM1, which promotes MT depolymerization, can act together to confer physiological dynamic properties upon purified tubulin³³. These proteins could regulate the length of spindle MTs, thereby contributing to the control of steady-state mitotic spindle

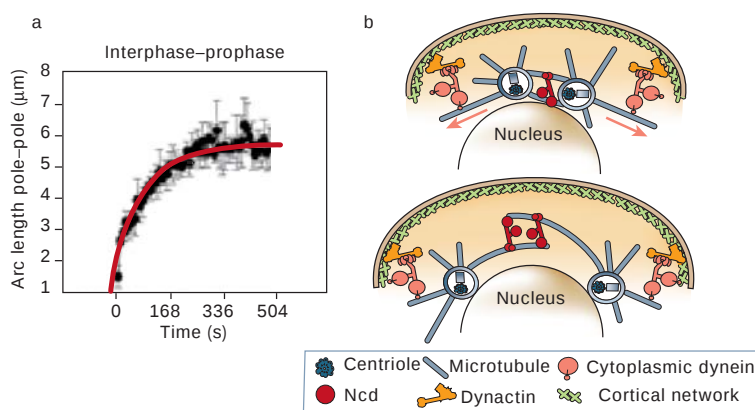
Box 1

Force-balance model for spindle pole motility

Motility in the spindle depends upon force generation by dynamic MT polymers and kinesin and dynein motors. MTs polymerize or depolymerize to generate pushing or pulling forces on an attached object such as a chromosome or spindle pole. Kinesin and dynein motors can step along the polymer lattice to exert forces on such a load. Simple physical arguments and, in some cases, experimental measurements suggest that the forces generated by MT dynamics and motor protein action are within the range of one to a few tens of piconewtons (refs 7,14). Because the spindle is capable of generating forces in the range of 1,000 pN (ref. 15), the question becomes: how do multiple force-generating elements (tens to thousands) cooperate as ensembles to generate forces and movements characteristic of the spindle?

Insights into this question emerge from a consideration of spindle pole motility during spindle morphogenesis in *Drosophila* embryos; in this system, bursts of spindle pole separation are punctuated by periods of stasis when pole–pole spacing is constant (isometric)^{4,27}. For example, plots of spindle pole separation (*S*) versus time (*t*) during early (interphase–prophase) mitosis (panel **a** in the figure above) reveal a burst of rapid motility that gradually slows to a stop at a constant (isometric) spacing of 6 μm . This spindle pole motility is proposed to depend upon a balance of opposing forces, as proposed for chromosome motility by Ostergren³⁶, such that outward forces (F_{out}) drive pole–pole separation, while inward forces (F_{in}) draw the poles together^{4,27,37}. In this model, the periods of stasis (for example, the 6- μm isometric state) are ‘equilibrium points’ when inward and outward forces balance one another.

It is proposed that the outward force during prophase, F_{out} , results from cortical dynein pulling on astral MTs (F_{dyn}) plus polymerizing MTs nucleated from one centrosome exerting pushing forces on the opposite centrosome (F_{pol}). The opposing inward force, F_{in} , is due to the minus-end-directed C-terminal kinesin Ncd acting on interpolar MTs to draw the poles together (F_{ncd})^{4,37}. Therefore, the net force driving spindle pole separation is $F(S) = F_{\text{out}} - F_{\text{in}} = \{F_{\text{dyn}}(S) + F_{\text{pol}}(S) - F_{\text{ncd}}(S)\}$. Intracellular motility events occur under conditions of low Reynolds number^{7,13,14,79}, and consequently the net force is proportional to the velocity^{7,79}. Thus, an equation of motion describing the dynamics of spindle pole motility is: $dS/dt = (1/\mu)\{F_{\text{dyn}}(S) + F_{\text{pol}}(S) - F_{\text{ncd}}(S)\}$, where μ is the effective drag coefficient at the nucleus, and dS/dt is the rate of spindle pole separation³⁷. When reasonable parameters and the geometry of the cytoplasm are used, a solution of this equation produces a reasonable fit to the experimental data³⁷ (in panel **a**; in the figure above, experimental data are shown in black, with the theoretical curve in red).



length³⁴, and indeed, a balance of activity between XMAP215 and XKCM1 in yeast is required for proper pole–pole separation during anaphase B³⁵. Thus, it seems likely that MAPs that control MT polymer dynamics, together with motors that crosslink and slide ipMTs outwards or inwards, could exert forces on spindle poles that control spindle length (Fig. 2c), as pole–pole spacing increases during spindle morphogenesis and elongation, for example^{4,27}.

The separation of the spindle poles that accompanies the morphogenesis and elongation of the spindle is an example of mitotic motility that reveals some of the basic principles by which cytoskeletal force generators drive the motility events underlying spindle mechanics (Box 1). Ostergren³⁶ proposed that shifts in a balance of antagonistic forces serve to move and position structures in the spindle, and evidence has accumulated showing that forces generated by growing and shrinking MTs and by antagonistic mitotic motors provide a molecular explanation for such a balance^{2,4,6,27,28}. Indeed a quantitative model (Box 1) can explain how a balance of opposing forces generated by ensembles of dynamic MTs and mitotic motors drives spindle pole motility^{4,27,37}, and similar models are likely to be relevant to other forms of motility (for example, chromosome motility).

Chromosome motility

During mitosis, pairs of sister chromatids associate with the spindle (Fig. 1b), congress to the spindle equator (Fig. 1c), and then segregate to opposite spindle poles (Fig. 1d, e)^{8,9}.

In the centrosome-directed spindle-assembly pathway, spindle morphogenesis and initial chromosome attachment are coupled, but in the centrosome-nucleated pathway, centrosome-nucleated MTs display dynamic instability, growing and shrinking in an exploratory

fashion to capture chromosomes in a timescale of minutes³⁸. Captured chromosomes often attach initially to the wall of spindle MTs by one sister kinetochore and move rapidly polewards at rates of 0.1 $\mu\text{m s}^{-1}$, possibly using dynein³⁹, before assuming a bi-oriented end-on configuration.

Once chromosomes have become bi-oriented, they congress to the equator to assume the metaphase configuration (Fig. 1b, c). During congression, bi-oriented chromosomes display ‘directional instability’, oscillating back-and-forth as episodes of poleward (P) motion and antipoleward (AP) motion at constant velocity are punctuated by rapid reversals, with the frequency of the reversals being biased so as to bring the chromosomes to the equator⁴⁰ (Fig. 2d). This bidirectional chromosome motility requires the elongation and shortening of kinetochore MTs (kMTs) and consequently much attention has focused on the important role played by MT dynamics⁹. However, the inhibition of motors such as dynein, chromokinesins and CENP-E can interfere with chromosome alignment, so motors must have some role^{41,42} (although the precise role of motors such as dynein in generating P forces remains controversial^{27,41,43}).

It is proposed that the bidirectional motility of chromosomes involves the integration of antagonistic P forces and AP forces acting along the pole–pole axis^{8,9,44} (Fig. 2e). P forces are likely to depend on the functional coordination of poleward MT flux, the depolymerization of MTs at kinetochores, minus-end-directed kinetochore–dynein motors, and kinetochore motors that couple kinetochore motility to MT dynamics. AP forces depend upon chromokinesins that push chromosome arms towards the spindle equator and generate a force gradient that diminishes with increasing distance from the spindle pole (Fig. 2b, d, e)^{8,9,40–42,44–47}. These opposing P and AP forces could produce tension in the kinetochore, the

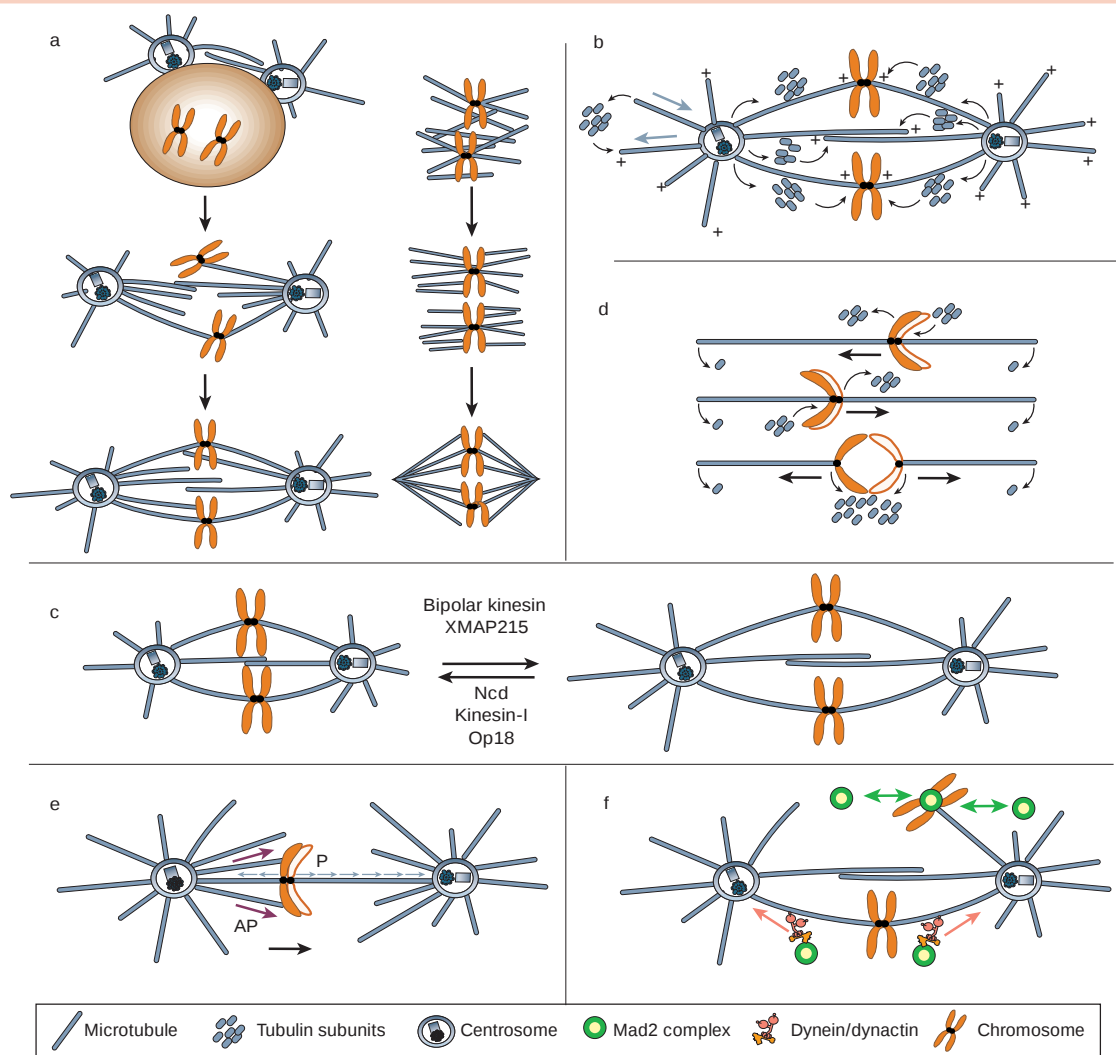


Figure 2 Spindle behaviour. **a**, In some systems, spindles assemble by a centrosome-directed pathway⁴ and in others by a chromosome-directed pathway^{5,6}. In the first case (left), two duplicated centrosomes nucleate the assembly of MTs, which produce a balance of outward and inward forces that drive pole–pole separation^{4,27,37}. In the latter pathway (right), condensed chromosomes direct the assembly of MTs, which are ‘sorted’ into a bipolar array by MT sliding motors and are crosslinked at their minus ends to form focused poles. **b**, Spindle MT dynamics coupled to GTP hydrolysis. Astral MTs grow and shrink by subunit addition or loss at their plus ends (dynamic instability). Kinetochore MTs (kMTs) and interpolar MTs (ipMTs) display poleward flux owing to subunit addition at their plus ends and loss at their minus ends, coupled with motor-dependent poleward translocation. Flux may exert poleward forces, F , that pull sister kinetochores towards opposite poles, where F is proportional to distance S from a pole³⁶. **c**, Control of spindle length by MT-associated proteins (MAPs) and motors. XMAP215 is a MAP that pushes the poles apart by promoting MT polymerization, whereas the kinesin-I motor XKCM1, and the MAP Op18 depolymerize MTs and shorten spindles. Bipolar kinesins crosslink and slide ipMTs outwards, thus lengthening the spindle, whereas Ncd, a C-terminal kinesin motor, slides ipMTs inwards. **d**, Kinetochore motility and MT dynamics. Prior to anaphase, bi-oriented chromosomes display kinetochore directional instability⁴⁰ in which kinetochore movement towards its facing pole (thick arrow) is coupled to MT disassembly at the leading kinetochore (upper panel). The kinetochore abruptly reverses its direction of movement, which becomes coupled to MT polymerization at the now trailing kinetochore (centre). During anaphase A, chromatid-to-pole motility is coupled to MT disassembly at the kinetochore and pole (lower panel). **e**, Shifts in a balance of pole-directed (P) and polar ejection (AP) forces control chromosome position. In one model⁴⁴ the ejection force within a half spindle forms a gradient that decays with distance from the pole and is opposed by the poleward force acting on the kinetochore. As a kinetochore is pulled polewards, the increasing AP force acts on chromosome arms to increase kinetochore tension, increasing the probability of MT dissociation from the leading kinetochore. Above a maximal tension, the leading kinetochore loses all its kMTs and, as a consequence, there is an abrupt reversal of direction as the sister kinetochore takes the lead. **f**, The spindle-assembly checkpoint. Once chromosomes are held under bipolar tension, dynein turns off the checkpoint by transporting Mad2 from kinetochores to poles.

magnitude of which increases as the chromosome moves polewards until, above a certain maximal tension, the direction of chromosome motion reverses abruptly; modulation of the frequency of reversals would then allow a chromosome to find the equator^{8,44}.

Anaphase is initiated after the bi-oriented chromosomes have aligned at the metaphase spindle equator, whereupon the links between sister chromatids dissolve and the separated sisters move to

opposite poles. Prior to the onset of anaphase, chromosomes are held at the equator by cohesin-mediated sister-chromatid cohesion and by chromokinesin-generated AP forces that push chromosome arms towards the equator^{48,49}. The cell-cycle dependent proteolysis of cohesin complexes allows the separation of the sister chromatids, while the degradation of the chromokinesin downregulates the AP forces, allowing the separated chromatids to be transported to

opposite spindle poles at speeds of $0.01\text{--}0.1\ \mu\text{m s}^{-1}$ (refs 8,9,40,41,43). A small force of magnitude $\sim 1\text{ pN}$ is sufficient to move a chromosome at these speeds⁷, although spindles are capable of exerting forces that are a few orders of magnitude greater than this^{7,15,50}, suggesting that tens to thousands of cytoskeletal force generators must be able to act cooperatively to generate the maximum forces acting during anaphase. Presumably the generation of the required P forces involves the functional coordination of minus-end-directed kinetochore motors, MT depolymerization at kinetochores and poleward MT flux^{27,41,43}.

Regulation of mitotic progression

Mitotic force generators located at the kinetochores do more than simply move and position chromosomes in the spindle, for they are also components of the spindle-assembly checkpoint, which delays the transition from metaphase to anaphase until all chromosomes are correctly aligned at the metaphase spindle equator (Fig. 1c). To do this, the checkpoint needs to detect a single unattached kinetochore among several properly attached ones.

Sister chromatid separation and exit from mitosis are controlled by the anaphase-promoting complex (APC), a ubiquitin-protein ligase that targets key proteins for proteolysis, for example, the cohesins and chromokinesins, whose destruction facilitates chromatid-to-pole motility. The checkpoint uses a network of proteins to inhibit the activity of the APC until all chromosomes are properly aligned, at which point the checkpoint is silenced allowing the APC to promote anaphase onset. Silencing depends upon the attachment of MTs to kinetochores, the formation of kMTs, and the establishment of bipolar tension at kinetochores^{10,51}.

A key player is the checkpoint protein Mad2, which is activated as a result of binding transiently to unattached kinetochores⁵², and is released as an active inhibitor of APC activity. Unattached kinetochores contain many other checkpoint proteins as well as tension-sensitive phospho-epitopes⁵³. Proper bipolar attachment of kinetochores leads to their dephosphorylation, and a dramatic relocalization of Mad2 (ref. 54). Transient kinetochore binding by Mad2 creates high steady-state levels of the protein on unattached kinetochores, but when kinetochores become properly aligned, Mad2 is depleted by translocation along kMTs to the spindle poles by the minus-end-directed dynein–dynactin complex⁵⁵ (Fig. 2f). Even a single unattached kinetochore can yield sufficient active Mad2 to inhibit the APC, but once they are all properly attached, Mad2 is sufficiently depleted to silence the checkpoint.

It should be emphasized that the kinetochore is a macromolecular complex containing tens, or possibly hundreds, of polypeptides, several of which may be involved in the spindle-assembly checkpoint. For example, ZW10 and Rod are required for correct targeting of dynein–dynactin to kinetochores and for checkpoint activation^{56,57}, although their precise role is unclear. In the absence of these proteins, the checkpoint is not activated even though high levels of Mad2 persist on unattached kinetochores, leading to proposals that they may normally serve to release activated Mad2 (ref. 10).

The CENP-E motor is proposed to be a mechanosensor of kinetochore tension, based on the observation that its depletion leads to a failure of checkpoint activation⁵⁸. However, it should be noted that it is difficult to decipher the precise function of CENP-E and other kinetochore motors during anaphase A, because they may participate directly in chromosome motility and also as components of the checkpoint regulatory system^{41,42,55,58}. Another fascinating area of uncertainty concerns whether the checkpoint detects MT attachment to kinetochores, tension on the kinetochore, or both. This issue is difficult to resolve experimentally because MT attachment and tension are inter-related — MT attachment leads to tension and tension can lead to more stable MT attachments⁵⁹, although only tension can discriminate between monopolar and bipolar attachment⁵¹. Accordingly, different experiments have led to different conclusions and further work is required^{54,59–62}.

The spindle-assembly checkpoint acts during metaphase when the spindle is maintained in an isometric state. In *Drosophila* embryonic spindles, isometric steady-state structures are also maintained by a balance of forces acting on astral and ipMTs at prophase, prometaphase and telophase⁴. We speculate that, as in metaphase, these three periods of stasis also allow the spindle to assess its mechanical status and to provide a stable framework to support MT-dependent breakdown of the nuclear envelope⁶³, chromosome capture and correct nuclear spacing, respectively.

How does the spindle determine the site of cytokinesis?

Cytokinesis, the final stage of division, creates two daughter cells from one parent cell (Fig. 1d–f). The position of the mitotic spindle during anaphase determines the location of the furrow, raising the question, what positions the spindle? Normally the spindle lies at the cell centre with its pole–pole axis parallel to the long axis of the cell, but sometimes the spindle is asymmetrically positioned leading to developmentally important asymmetric divisions. It is plausible that a balance of forces is responsible for cell centring of the spindle and that a shift in this balance leads to asymmetric spindle positioning^{64,65}.

Spindle MTs determine the position of the cleavage plane midway between the poles (Fig. 1d), and early micromanipulation experiments suggested that the astral MTs are responsible for inducing the cleavage furrow⁶⁶ with the corresponding signal being proportional to the number of astral MTs reaching the cortex. But other data suggest that ipMTs and kMTs, rather than astral MTs, are responsible for furrow positioning⁶⁷. Some evidence suggests that chromosomes play a role, because some proteins — called chromosome passenger proteins — translocate from chromosomes to spindle MTs, and could therefore control the position of the furrow. Thus, it seems that the whole spindle mediates furrow induction, but there are system-specific differences in the parts of the spindle that are important⁶⁸.

The initiation of cytokinesis begins a few minutes after anaphase onset. There is no obvious tight temporal coupling between the completion of mitosis and the beginning of cytokinesis, but there seems to exist a permissive time interval lasting a few minutes after mitotic exit when cytokinesis can occur. In echinoderm eggs, it has been estimated⁶⁶ that the signalling event occurs approximately 5 min prior to furrow formation, taking about a minute for the signal to travel from the asters to the furrow, and a further 2.5 min for the furrow to develop. These timescales suggest that motor-mediated transport of signalling molecules is involved. For example, a motor protein moving at $\sim 0.1\ \mu\text{m s}^{-1}$ would travel $\sim 6\ \mu\text{m}$ over 1 min, which is roughly the thickness of the egg cortex.

It is tempting to speculate that the delivery of this signal as well as other spindle-associated transport events discussed below depends upon a two-step transport system involving kinesin-driven motility along astral MTs followed by myosin-driven motility through cortical actin (Fig. 1d inset). In this way it would be analogous to the pathway of vesicle recruitment for exocytosis during cell membrane resealing and neurotransmission⁶⁹. Elegant experiments have focused on a class of vesicles that seem to be delivered by astral MTs specifically to the contractile ring⁷⁰, although this is a late event in cytokinesis. Other factors that may be involved in the spatiotemporal control of furrow initiation include the small GTPase RhoA, and cell cycle-regulated myosin light-chain kinases, which may contribute to the timing of cytokinesis⁷¹. Functional proteomic approaches to cytokinesis^{12,72} may uncover the molecules involved in this and other aspects of cytokinesis.

Once the division plane has been established, the assembly of the actomyosin-based contractile ring is crucial for subsequent ingression (Fig. 1e and inset). The accumulation of actin and myosin II in the region of the furrow occurs during late anaphase by an uncertain mechanism. Some pre-existing actin filaments are recruited into the cleavage furrow by directed transport in the plane of the cortex, probably powered by myosin⁶⁷, but additional actin

polymerization, as well as recruitment of myosin filaments from the underlying cytoplasm, may occur as well.

Force generation for furrow ingression

The nature of the mechanical process that underlies ingression has been a topic of intense debate, specifically whether relaxation at the poles of the cell or contraction at the equator is responsible. The myosin-dependent equatorial contraction model (Fig. 1e) has prevailed^{73–75}. Structural studies showed an abundance of organized actomyosin bundles in the contractile ring aligned along the cell equator consistent with a 'purse-string' sliding-filament mechanism. It is estimated that hundreds to thousands of myosin molecules must be localized within this structure where they cooperate to generate the maximal contractile force of 10^3 to 10^5 pN that is proposed to be developed by the contractile ring^{16,17,66,75}. However, there are reasons to question the generality of the equatorial contraction model, because highly ordered actomyosin bundles are not a universal feature of the contractile ring, and perhaps contraction could be spread globally throughout the cortex⁷⁵, rather than being restricted to the ring.

During ingression, the cell membrane deforms owing to its attachments with an underlying actomyosin network, and consequently ingression is often accompanied by the fusion of membrane vesicles with the ingressing cell membrane behind the leading edge of the furrow^{70,76}. This targeted vesicle insertion contributes to narrowing the distance between the tip of the furrow and the spindle mid-zone, in addition to supplying proteins and lipids. It is plausible that a motor transport system may move the vesicles along spindle MTs and cortical actin filaments to the cell surface⁶⁹ for fusion with the plasma membrane^{70,76}.

Completion of cytokinesis

Conventional models of cytokinesis posit that once the furrow is positioned, ingression proceeds independently of the spindle or other MT structures by means of the self-enhancing contractility of the actomyosin network. There is growing evidence, however, of a dynamic interplay between the spindle and the actomyosin cortex: successful completion of cytokinesis requires a host of proteins of uncertain function that localize to the central spindle, including MT-based motors, septins, formin-homology proteins, components of the telophase disc, kinases and GTPases and their regulators^{14,71,74}. Among these, a kinesin motor (variously named MKLP1/CHO1/PavKLP/ZEN4) is perhaps best understood as it is crucial in organizing the arrays of antiparallel MTs that form the spindle mid-zone and, as a consequence, it is essential for the completion of cytokinesis¹¹.

Ingression continues until the contractile ring compresses the central spindle into a compact midbody containing an electron-dense 'matrix' at its midline (Fig. 1f). It is proposed that midbody MTs could then serve as tracks for the motor-mediated transport of Golgi-derived vesicles and signalling molecules to the furrow during cell–cell abscission (Fig. 1f)^{70,76}. The reasoning behind this is that, during the process of abscission, the midbody is severed and a secreted membrane barrier partitions the cells to form two daughter cells in a process that is thought to resemble the recruitment of vesicles for Ca^{2+} -regulated exocytosis⁶⁹ (Fig. 1f inset).

Future work

The prodigious amount of high quality work, much of which could not be cited, bears witness to our persistent fascination with cell division. For example, we did not mention work on plant cell division, where studies of the phragmoplast have pioneered thinking about cell–cell abscission⁷⁷, or work on bacterial cell division, where a tubulin-like protein, FtsZ and an actin-like protein, Par-M seem to drive cytokinesis and plasmid segregation, respectively⁷⁸, in an apparent reversal of the eukaryotic paradigm. Despite this effort, much needs to be done to understand mitosis and cytokinesis. The

important role that cytoskeletal proteins play as force-generating elements in these processes is clear. The identities of many players are known^{4–6,10,13}, and there exists a reasonable understanding of the physical mechanisms by which polymer dynamics and motor proteins generate force for motility at the individual level^{7,14}. But how these force-generating elements function cooperatively within the ensembles that form the cell division machinery is much less clear^{7,37}, and improved understanding will require quantitative biophysical and biochemical analysis combined with theoretical modelling in normal and experimentally manipulated dividing cells. These are exciting times for students of mitosis and cell division. □

doi:10.1038/nature01599

- Harris, H. *The Birth of the Cell* (Yale Univ. Press, New Haven, 1999).
- Mitchison, T. J. & Salmon, E. D. Mitosis: a history of division. *Nature Cell Biol.* **3**, E17–E22 (2001).
- Wood, K. W., Cornwell, W. D. & Jackson, J. R. Past and future of the mitotic spindle as an oncology target. *Curr. Opin. Pharmacol.* **1**, 370–377 (2001).
- Sharp, D. J., Rogers, G. C. & Scholey, J. M. Microtubule motors in mitosis. *Nature* **407**, 41–47 (2000).
- Karsenti, E. & Vernos, I. The mitotic spindle, a self-made machine. *Science* **294**, 543–547 (2001).
- Wittman, T., Hyman, A. & Desai, A. The spindle, a dynamic assembly of microtubules and motors. *Nature Cell Biol.* **3**, E28–E34 (2001).
- Scholey, J. M. & Mogilner, A. in *Molecular Motors* (ed. Schliwa, M.) 327–355 (Wiley-VCH, Weinheim, 2003).
- Kapoor, T. M. & Compton, D. A. Searching for the middle ground: mechanisms of chromosome alignment during mitosis. *J. Cell Biol.* **157**, 551–556 (2002).
- McIntosh, J. R., Grishunk, E. L. & West, R. R. Chromosome-microtubule interactions during mitosis. *Annu. Rev. Cell Dev. Biol.* **18**, 193–219 (2002).
- Shah, J. V. & Cleveland, D. W. Waiting for anaphase: Mad2 and the spindle assembly checkpoint. *Cell* **103**, 997–1000 (2000).
- Glotzer, M. Animal cell cytokinesis. *Annu. Rev. Cell Dev. Biol.* **17**, 351–386 (2001).
- Gonczy, P. et al. Functional genomic analysis of cell division in *C. elegans* using RNAi of genes on chromosome III. *Nature* **408**, 331–336 (2000).
- Bray, D. *Cell Movements* (Garland, New York, 2001).
- Howard, J. *Mechanics of Motor Proteins and the Cytoskeleton* (Sinauer, Sunderland, MA, 2001).
- Nicklas, R. B. Measurements of the force produced by the mitotic spindle in anaphase. *J. Cell Biol.* **97**, 542–548 (1983).
- Rappaport, R. Cell division: direct measurement of maximum tension exerted by furrow of echinoderm eggs. *Science* **156**, 1241–1243 (1967).
- Burton, K. & Taylor, D. L. Traction forces of cytokinesis measured with optically modified elastic substrata. *Nature* **385**, 450–454 (1997).
- Gunawardane, R. N., Lizarraga, S. B., Wiese, C., Wilde, A. & Zheng, Y. Gamma-tubulin complexes and their role in microtubule nucleation. *Curr. Top. Dev. Biol.* **49**, 55–73 (2000).
- Moritz, M., Braundfeld, M. B., Guenebaut, V., Heuser, J. & Agard, D. A. Structure of the γ -tubulin ring complex: a template for microtubule nucleation. *Nature Cell Biol.* **2**, 365–370 (2000).
- Keating, T. J. & Borisov, G. G. Immunofluorescence evidence for the template mechanism of microtubule nucleation. *Nature Cell Biol.* **2**, 352–357 (2000).
- Erickson, H. P. γ -tubulin nucleation: template or protofilament? *Nature Cell Biol.* **2**, E93–E96 (2000).
- Popov, A. V., Severin, F. & Karsenti, E. XMAP215 is required for the microtubule-nucleating activity of centrosomes. *Curr. Biol.* **12**, 1326–1330 (2002).
- Kalab, P., Weis, K. & Heald, R. Visualization of a Ran-GTP gradient in interphase and mitotic *Xenopus* egg extracts. *Science* **295**, 2452–2456 (2002).
- Walczak, C. E. Ran hits the ground running. *Nature Cell Biol.* **3**, E1–E3 (2001).
- Theurkauf, W. E. TACCing down the spindle poles. *Nature Cell Biol.* **3**, E159–E161 (2001).
- Walczak, C. E., Vernos, I., Mitchison, T. J., Karsenti, E. & Heald, R. Model for the proposed roles of different microtubule based motor proteins in establishing spindle bipolarity. *Curr. Biol.* **8**, 903 (1998).
- Brust-Mascher, I. & Scholey, J. M. Microtubule flux and sliding in mitotic spindles of early *Drosophila* embryos. *Mol. Biol. Cell* **13**, 3967–3975 (2002).
- Nedelec, F. Computer simulations reveal motor properties generating stable antiparallel microtubule interactions. *J. Cell Biol.* **158**, 1005–1015 (2002).
- Kapoor, T. M. & Mitchison, T. J. Eg5 is static in bipolar spindles relative to tubulin: evidence for a static spindle matrix. *J. Cell Biol.* **154**, 1125–1133 (2001).
- Wells, W. A. Searching for a spindle matrix. *J. Cell Biol.* **154**, 1102–1104 (2001).
- Wigge, P. A. et al. Analysis of the *Saccharomyces* spindle pole by matrix assisted laser desorption/ionization (MALDI) mass spectrometry. *J. Cell Biol.* **141**, 967–977 (1998).
- Mack, G. J. & Compton, D. A. Analysis of mitotic microtubule-associated proteins using mass spectrometry identifies astrin, a spindle-associated protein. *Proc. Natl Acad. Sci. USA* **98**, 14434–14439 (2001).
- Kinoshita, K., Arnal, I., Desai, A., Drechsel, D. N. & Hyman, A. A. Reconstitution of physiological microtubule dynamics using purified components. *Science* **294**, 1340–1343 (2001).
- Heald, R. A dynamic duo of microtubule regulators. *Nature Cell Biol.* **2**, E11–E12 (1999).
- Severin, F., Haberman, B., Huffaker, T. & Hyman, A. A. Stu2 promotes mitotic spindle elongation in anaphase. *J. Cell Biol.* **153**, 435–442 (2001).
- Ostergren, G. The mechanism of co-orientation in bivalents and multivalents. The theory of orientation by pulling. *Hereditas* **37**, 85–156 (1951).
- Cytrynbaum, E., Scholey, J. M. & Mogilner, A. Force-balance model for early spindle pole separation in *Drosophila* embryonic mitotic spindles. *Biophys. J.* **84**, 757–769 (2003).
- Holy, T. E. & Leibler, S. Dynamic instability of microtubules as an efficient way to search in space. *Proc. Natl Acad. Sci. USA* **91**, 5682–5685 (1994).
- Alexander, S. P. & Rieder, C. L. Chromosome motion during attachment to the vertebrate spindle: initial saltatory-like behavior of chromosomes and quantitative analysis of force production by

- nascent kinetochore fibers. *J. Cell Biol.* **113**, 805–815 (1991).
40. Skibbens, R. V., Skeen, V. P. & Salmon, E. D. Directional instability of kinetochore motility during chromosome congression and segregation in mitotic newt lung cells: a push-pull mechanism. *J. Cell Biol.* **122**, 859–875 (1993).
41. Sharp, D. J., Rogers, G. C. & Scholey, J. M. Cytoplasmic dynein is required for poleward chromosome movement in *Drosophila* embryos. *Nature Cell Biol.* **2**, 922–930 (2000).
42. Yucel, J. K. *et al.* CENP-meta, an essential kinetochore kinesin required for the maintenance of metaphase chromosome alignment in *Drosophila*. *J. Cell Biol.* **150**, 1–12 (2000).
43. Maddox, P., Desai, A., Oegema, K., Mitchison, T. J. & Salmon, E. D. Poleward microtubule flux is a major component of spindle dynamics and anaphase A in mitotic *Drosophila* embryos. *Curr. Biol.* **12**, 1670–1674 (2002).
44. Joglekar, A. P. & Hunt, A. J. A simple, mechanistic model for directional instability during mitotic chromosome movements. *Biophys. J.* **83**, 42–58 (2002).
45. Levesque, A. A. & Compton, D. A. The chromokinesin Kid is necessary for chromosome arm orientation and oscillation, but not congression, on mitotic spindles. *J. Cell Biol.* **154**, 1135–46 (2001).
46. Maney, T., Ginkel, L. M., Hunter, A. W. & Wordeman, L. The kinetochore of higher eucaryotes: a molecular view. *Int. Rev. Cytol.* **194**, 67–131 (2000).
47. Desai, A., Verma, S., Mitchison, T. J. & Walczak, C. E. Kin-I kinesins are microtubule-destabilizing enzymes. *Cell* **96**, 69–78 (1999).
48. Nasmyth, K., Peters, J. M. & Uhlmann, F. Splitting the chromosome: cutting the ties that bind sister chromatids. *Science* **288**, 1379–1384 (2000).
49. Funabiki, H. & Murray, A. W. The *Xenopus* chromokinesin, Xkid is essential for metaphase chromosome alignment and must be degraded to allow anaphase chromosome movement. *Cell* **102**, 411–424 (2000).
50. Marshall, W. F., Marko, J. F., Agard, D. A. & Sedat, J. W. Chromosome elasticity and mitotic polar ejection force measured in living *Drosophila* embryos by four-dimensional microscopy-based motion analysis. *Curr. Biol.* **11**, 569–578 (2001).
51. Zhou, J., Yao, J. & Yoshi, H. C. Attachment and tension in the spindle assembly checkpoint. *J. Cell Sci.* **115**, 3547–3555 (2002).
52. Howell, B. J., Hoffman, D. B., Fang, G., Murray, A. W. & Salmon, E. D. Visualization of Mad2 dynamics at kinetochores, along spindle fibers, and at spindle poles in living cells. *J. Cell Biol.* **150**, 1233–1250 (2000).
53. Li, X. & Nicklas, R. B. Tension-sensitive kinetochore phosphorylation and the chromosome distribution checkpoint in praying mantid spermatocytes. *J. Cell Sci.* **110**, 537–545 (1997).
54. Hoffman, D. B., Pearson, C. G., Yen, T. J., Howell, B. J. & Salmon, E. D. Microtubule-dependent changes in assembly of microtubule motor proteins and mitotic spindle checkpoint proteins at PtK1 kinetochores. *Mol. Biol. Cell* **12**, 1995–2009 (2001).
55. Howell, B. J. *et al.* Cytoplasmic dynein/dynactin drives kinetochore protein transport to the spindle poles and has a role in mitotic spindle checkpoint inactivation. *J. Cell Biol.* **155**, 1159–1172 (2001).
56. Chan, G. K. T., Jablonski, S. A., Starr, D. A., Goldberg, M. L. & Yen, T. J. Human Zw10 and ROD are mitotic checkpoint proteins that bind to kinetochores. *Nature Cell Biol.* **2**, 944–947 (2000).
57. Wójcik, E. *et al.* Kinetochore dynein: its dynamics and role in the transport of the Rough deal checkpoint protein. *Nature Cell Biol.* **3**, 1001–1008 (2001).
58. Abrieu, A., Kahana, J. A., Wood, K. W. & Cleveland, D. W. Cenp-E as an essential component of the mitotic checkpoint in vitro. *Cell* **102**, 817–826 (2000).
59. Nicklas, R. B., Waters, J. C., Salmon, E. D. & Ward, S. C. Checkpoint signals in grasshopper meiosis are sensitive to microtubule attachment but tension is still essential. *J. Cell Sci.* **114**, 4173–4183 (2001).
60. Skoufias, D. A., Andreassen, P. R., Lacroix, F. B., Wilson, L. & Margolis, R. L. Mammalian mad2 and bub1/bubR1 recognize distinct spindle-attachment and kinetochore-tension checkpoints. *Proc. Natl Acad. Sci. USA* **98**, 4492–4497 (2001).
61. Shannon, K. B., Canman, J. C. & Salmon, E. D. Mad2 and BubR1 function in a single checkpoint pathway that responds to a loss of tension. *Mol. Biol. Cell* **10**, 1091/mbc.E02-03-0137 (2002).
62. Zhou, J., Panda, D., Landen, J. W., Wilson, L. & Joshi, H. C. Minor alteration of microtubule dynamics causes loss of tension across kinetochore pairs and activates the spindle checkpoint. *J. Biol. Chem.* **277**, 17200–17208 (2002).
63. Beaudouin, J., Gerlich, D., Daigle, N., Eils, R. & Ellenberg, J. Nuclear envelope breakdown proceeds by microtubule-induced tearing of the lamina. *Cell* **108**, 83–96 (2002).
64. Reinsch, S. & Gonczy, P. Mechanisms of nuclear positioning. *J. Cell Sci.* **111**, 2283–2295 (1998).
65. Grill, S. W., Gonczy, P., Stelzer, E. H. & Hyman, A. Polarity controls forces governing asymmetric spindle positioning in the *C. elegans* embryo. *Nature* **409**, 630–633 (2001).
66. Rappaport, R. *Cytokinesis in Animal Cells* (Cambridge Univ. Press, Cambridge, 1996).
67. Cao, L. G. & Wang, Y. L. Mechanism of the formation of contractile ring in dividing cultured animal cells. I. Recruitment of preexisting actin filaments into the cleavage furrow. *J. Cell Biol.* **110**, 1089–1095 (1990).
68. Wang, Y. L. The mechanism of cytokinesis: reconsideration and reconciliation. *Cell Struct. Funct.* **26**, 633–638 (2001).
69. Bi, G. Q. *et al.* Kinesin- and myosin-driven steps of vesicle recruitment for Ca^{2+} regulated exocytosis. *J. Cell Biol.* **138**, 999–1008 (1997).
70. Shuster, C. B. & Burgess, D. R. Targeted new membrane addition in the cleavage furrow is a late, separate event in cytokinesis. *Proc. Natl Acad. Sci. USA* **99**, 3633–3638 (2002).
71. Robinson, D. N. & Spudich, J. A. Towards a molecular understanding of cytokinesis. *Trends Cell Biol.* **10**, 228–237 (2000).
72. Somma, M. P., Fasulo, B., Cenci, G., Cundari, E. & Gatti, M. Molecular dissection of cytokinesis by RNA interference in *Drosophila* cultured cells. *Mol. Biol. Cell* **13**, 2448–2460 (2002).
73. Satterwhite, L. L. & Pollard, T. D. Cytokinesis. *Curr. Opin. Cell Biol.* **4**, 43–52 (1992).
74. Guertin, D. A., Trautmann, S. & McCollum, D. Cytokinesis in eukaryotes. *Microbiol. Mol. Biol. Rev.* **66**, 155–178 (2002).
75. Robinson, D. N., Cavet, G., Warrick, H. M. & Spudich, J. A. Quantitation of the distribution and flux of myosin-II during cytokinesis. *BMC Cell Biol.* **3**, 4–16 (2002).
76. Finger, F. P. & White, J. G. Fusion and fission: membrane trafficking in animal cytokinesis. *Cell* **108**, 727–730 (2002).
77. Liu, B. & Lee, Y.-R. Kinesin-related proteins in plant cytokinesis. *J. Plant Growth Regul.* **20**, 141–150 (2001).
78. Moller-Jensen, J., Jensen, R. B., Lowe, J. & Gerdes, K. Prokaryotic DNA segregation by an actin-like filament. *EMBO J.* **21**, 3119–3127 (2002).
79. Purcell, E. Life at low Reynolds number. *Am. J. Phys.* **45**, 3–11 (1977).

Acknowledgements Supported by the National Institutes of Health.

Dynamics and mechanics of the microtubule plus end

Joe Howard & Anthony A. Hyman

Max Planck Institute of Molecular Cell Biology and Genetics (MPI-CBG), Pfotenhauerstrasse 108, 01307 Dresden, Germany
(e-mail: howard@mpi-cbg.de; hyman@mpi-cbg.de)

An important function of microtubules is to move cellular structures such as chromosomes, mitotic spindles and other organelles around inside cells. This is achieved by attaching the ends of microtubules to cellular structures; as the microtubules grow and shrink, the structures are pushed or pulled around the cell. How do the ends of microtubules couple to cellular structures, and how does this coupling regulate the stability and distribution of the microtubules? It is now clear that there are at least three properties of a microtubule end: it has alternate structures; it has a biochemical transition defined by GTP hydrolysis; and it forms a distinct target for the binding of specific proteins. These different properties can be unified by thinking of the microtubule as a molecular machine, which switches between growing and shrinking modes. Each mode is associated with a specific end structure on which end-binding proteins can assemble to modulate dynamics and couple the dynamic properties of microtubules to the movement of cellular structures.

The textbook functions of microtubules are to act as beams that provide mechanical support for the shape of cells, and as tracks along which molecular motors move organelles from one part of the cell to another (Fig. 1a). To perform these functions, a cell must control the assembly and orientation of its microtubule cytoskeleton. Microtubules assemble by polymerization of α - β dimers of tubulin. Polymerization is a polar process that reflects the polarity of the tubulin dimer, which in turn dictates the polarity of the microtubule (Fig. 2a). *In vitro*, purified tubulin polymerizes more quickly from the plus end, which is terminated by the β -subunit. The other, slow-growing end is known as the minus end, and is terminated by the α -subunit. In animal cells, minus ends are generally anchored at centrosomes, which consist of specialized microtubule-based structures called centrioles, surrounded by pericentriolar proteins¹ (Fig. 1b). In yeast, the analogous structure is the spindle pole body². An important component of the centrosome is an unusual form of tubulin, γ -tubulin, which is thought to initiate nucleation by forming rings that act as templates for new microtubule growth^{3,4}. After nucleation, microtubules grow out with their plus ends leading into the cytoplasm. Thus to a first approximation, the distribution of the microtubule cytoskeleton is determined by the location of the centrosome.

The first clue as to how cells rearrange the distribution of microtubules came from the discovery that during the polymerization of pure tubulin, plus ends switch between phases of slow growth and rapid shrinkage⁵ (Fig. 2b). The conversion from growing to shrinking is called catastrophe, whereas the conversion from shrinking to growing is called rescue (Fig. 2b). Analysis in tissue culture cells^{6,7} and in cellular extracts⁸ soon confirmed that this behaviour, termed dynamic instability, is a feature of microtubules growing under physiological conditions (for a review, see ref. 9).

The importance of the discovery of dynamic instability was that it provided for the first time a mechanism by which microtubules could reassemble into different structures during the cell cycle or during development. It was hypothesized

that microtubules could grow out and if they made productive interactions with cellular structures¹⁰ or soluble cues^{11,12}, they would be stabilized. An early confirmation of this idea was the finding that kinetochores, specialized structures that connect microtubules to chromosomes, can 'capture' and stabilize the end of a growing microtubule¹³. Recently, soluble cues have also been shown to modulate microtubule dynamics during spindle assembly in *Xenopus* egg extracts. Here a Ran-dependent signal changes the local environment of cytoplasm around the chromosomes, stabilizing the plus ends and initiating the assembly of the mitotic spindle (for a recent review, see ref. 14).

Microtubules as molecular machines

Once assembled, polarized arrays of microtubules provide tracks for the transport of organelles and chromosomes¹⁵. This transport is driven by motor proteins such as kinesin and dynein that interact with and move along the lateral surface of the microtubule. Motor proteins are molecular machines — they transduce chemical energy derived from ATP hydrolysis into mechanical work used for cellular motility — and there has been considerable interest recently in understanding the biophysical mechanisms by which these protein machines work^{16,17}.

But examples of cellular motility exist that do not rely exclusively on motor proteins. One is the movement of chromosomes during metaphase and anaphase of mitosis (Fig. 3a). After the plus ends of microtubules have attached to the chromosome via the kinetochore¹⁸, the growth and shrinkage of these kinetochore-attached microtubules move the chromosome away from or towards the pole to which the minus end of the microtubule is attached¹⁹. Other examples are provided by the movement of the nucleus or the mitotic spindle through interactions between microtubules and the cell cortex, where the cortex is loosely defined as the plasma membrane and its associated protein components. Such cortical interactions, inferred from experiments in embryonic systems such as *Caenorhabditis elegans* (Fig. 3b) or *Drosophila*^{20,21}, have now been viewed directly in yeast. In the fission yeast *Schizosaccharomyces pombe*, microtubules grow out from the spindle pole bodies

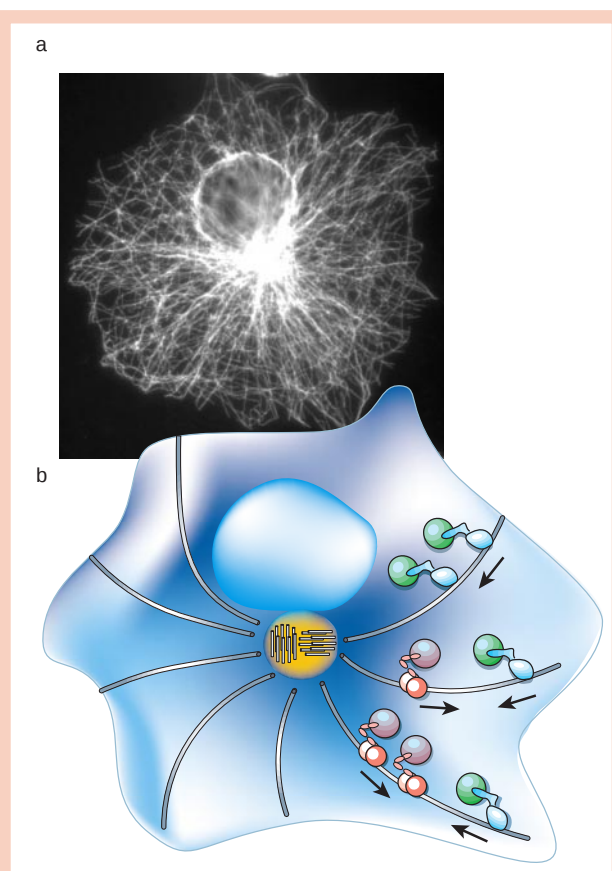


Figure 1 Microtubules are dynamic polymers. **a**, An interphase cell stained with an antibody to tubulin. Microtubules extend from the centrosome throughout the cell. (Image courtesy of A. Akhmanova.) **b**, A schematic diagram of the cell. Centrioles are shown in the centrosome (yellow). Red circles denote vesicles moving to the outside of the cell. Green circles denote vesicles moving to the centrosome.

and push back on the nucleus when their plus ends reach the ends of the cell²². The pushing from the two ends of the cell centres the nucleus. In the yeast *Saccharomyces cerevisiae*, cells divide by budding, resulting in a mother and a daughter cell. Prior to division, microtubules growing from one of the spindle pole bodies enter the bud where they attach to the cortex. The depolymerization of these

cortex-attached microtubules is thought to reel in the spindle so that one of the poles is now located in the bud and will be inherited by the daughter following division^{23–26} (Fig. 3c).

These examples suggest that microtubules themselves, in the absence of motors, can move cellular structures around inside cells by maintaining attachments as they grow or shrink¹⁹. *In vitro* studies with purified tubulin have confirmed that the end of a microtubule can act as a molecular machine that converts chemical energy into mechanical work, just like a motor protein. Polymerizing microtubules can deform membranes²⁷ or induce microtubule buckling²⁸, while depolymerizing microtubules can move beads attached to their ends²⁹. Furthermore, the forces generated are high — up to 4 pN — which indicates that microtubule dynamics can generate as much force as motor proteins¹⁶. These forces can be used to form structures *in vitro*. Indeed, if an aster of outward-growing microtubules is placed in a microfabricated chamber, the pushing forces are capable of centring the aster^{30,31}, analogous to the centring of the nucleus in yeast²². Thus the microtubule end can be thought of as a molecular machine. Because microtubules grow and shrink by addition and loss of subunits from their ends, coupling of microtubule pulling and pushing to mechanical work can be distilled to the problem of the nature and control of the plus end of the microtubule.

GTP hydrolysis cycle

The energy to drive the microtubule machine comes from GTP hydrolysis. Tubulin is a GTPase whose activity is stimulated by polymerization³². A crucial observation is that tubulin polymerizes in the presence of non-hydrolysable GTP to form stable microtubules³³. Thus, polymerization is driven by the high affinity of the tubulin–GTP dimer for the end of the microtubule. The high affinity means that polymerization will take place even against compressive forces, theoretically as high as several piconewtons¹⁶, accounting for the ability of a growing microtubule to do work. But the high stability of the GTP microtubule poses a problem for disassembly, because GTP microtubules depolymerize at a negligible rate and evidently cannot do work while shortening. This problem is solved by GTP hydrolysis. The resulting GDP microtubule is very unstable and, if allowed to, will depolymerize even in the presence of tensile forces that oppose the depolymerization. Thus, binding of the GTP subunit can do work during the growth phase while unbinding of the GDP subunit can do work during the shrinkage phase.

There are two key regulatory events in the GTP cycle. The first is the coupling of hydrolysis to polymerization (for a detailed discussion, see ref. 34). An elegant coupling mechanism has been provided by the determination of the atomic structure of tubulin (Fig. 4a). In a microtubule, the β -subunit resides at the plus end³⁵. The structure

Figure 2 Microtubule structure and dynamics.

a, A microtubule lattice. The β -subunit of tubulin is on the plus end. **b**, Dynamic instability of microtubules. Microtubules growing out from a centrosome switch between phases of growing and shrinking. The figure shows a hypothetical aster at two different times. The different colours represent different microtubules. The red and yellow microtubules are shrinking at both times. The blue microtubule is growing at both times. The green microtubule, growing at the first time, has undergone a catastrophe by the second time. The brown microtubule, shrinking at the first time, has undergone a rescue by the second time.

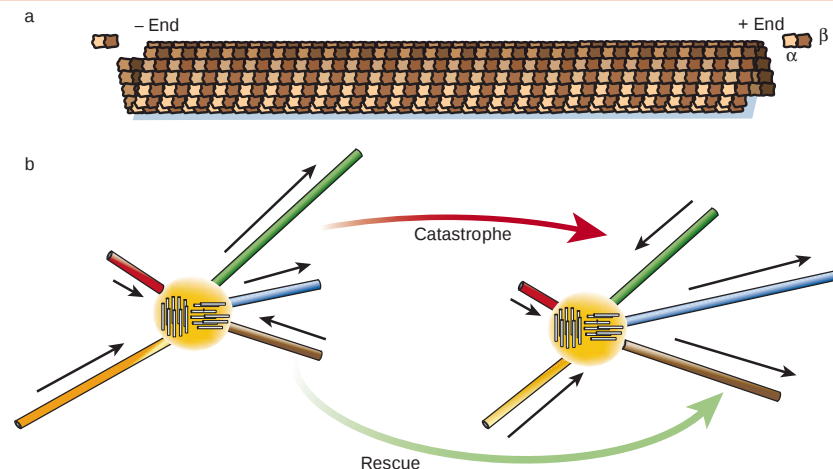
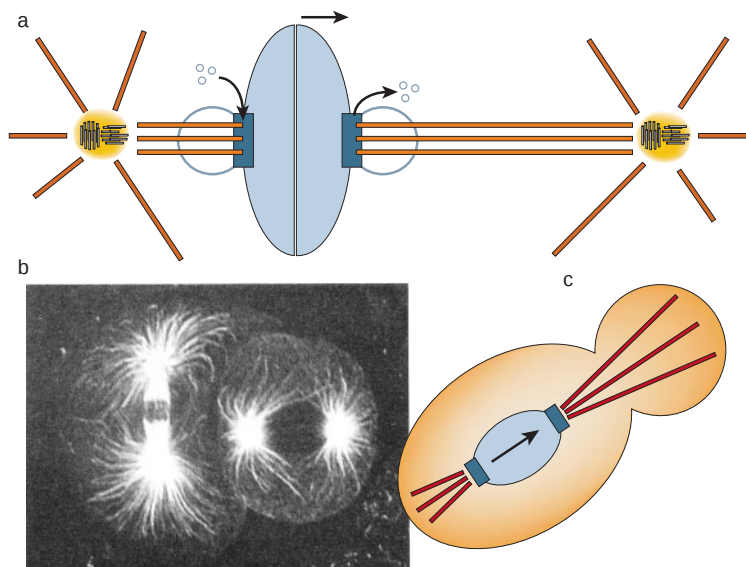


Figure 3 Interaction of microtubule ends with cellular structures. **a**, During metaphase of mitosis, movement of the chromosome (to the right) is associated with polymerization of microtubules on one side (left) and depolymerization on the other (right). **b**, Two-cell stage *Caenorhabditis elegans* embryo. One spindle (on the right) is rotated with respect to the other, perhaps through interactions between microtubules and a cortical site located between the two cells. **c**, Movement of the *Saccharomyces cerevisiae* spindle pole into the bud (at the right). Microtubules from one of the spindle pole bodies attach to the bud cortex. Depolymerization of these microtubules at the cortex may reel in the spindle into the bud.



shows that, although the β -subunit pocket can bind GTP, it lacks crucial residues necessary for hydrolysis. These residues are donated by the α -subunit when it docks to the end, and in this way hydrolysis is triggered³⁶ (Fig. 4b). If hydrolysis is faster than polymerization then the structural findings support a simple model in which a single ring of GTP subunits stabilizes the microtubule plus end by preventing internal GDP subunits from dissociating^{37,38}. On the other hand, if hydrolysis lags behind polymerization, then a large cap of GTP subunits may form at the end and this could further stabilize the microtubule. Removal of this cap and the triggering of microtubule depolymerization constitutes the second key regulatory event. But we know a lot less about this event than the coupling of hydrolysis to polymerization. Recent work on the structure of the microtubule end, and proteins that bind to the end, is beginning to shed light on this issue.

Structure of the microtubule end

If a microtubule end is to act as a molecular machine, then it must undergo conformational changes in response to GTP hydrolysis. For example, motor proteins undergo a structural transition, known as the powerstroke, that is driven by the ATP hydrolysis cycle and that leads to the generation of force and the production of mechanical work^{16,17,39}. Analogous changes do indeed take place at the ends of the microtubule. Viewing growing and shrinking microtubules in vitreous ice has shown that, both for pure tubulin and for microtubules growing under physiological conditions, the ends of growing microtubules (Fig. 4c) consist of two-dimensional sheets of protofilaments (head-to-tail arrangements of tubulin dimers)^{40,41}, whereas the ends of shrinking microtubules (Fig. 4d) are frayed, often resembling rams' horns^{41,42}. Therefore it seems clear that there is a structural transition associated with the switch between growing and shrinking.

How does GTP hydrolysis control this structural transition? The early discovery of protofilament rings as depolymerization products of microtubules led to the hypothesis that GTP hydrolysis destabilizes the lattice by increasing the curvature of the protofilament^{43,44}. Thus in the GTP state the subunits form straight protofilaments that fit nicely into the wall of the microtubules, whereas in the GDP state they form bent protofilaments that want to splay out from the lattice (Fig. 4d). Recent work has provided strong additional evidence for this model. First, protofilaments made from GTP-tubulin are straighter than those made from GDP-tubulin⁴⁵. Second, the

structure of the tubulin-sequestering protein Op18/stathmin complexed with two tubulin-GDP dimers shows the dimers are bent⁴⁶. Although we do not know whether the bend is introduced by Op18 or not, it is suggestive that the bend within the dimer, together with rotation between the dimers, generates a protofilament with the same curvature as a GDP protofilament measured by other means.

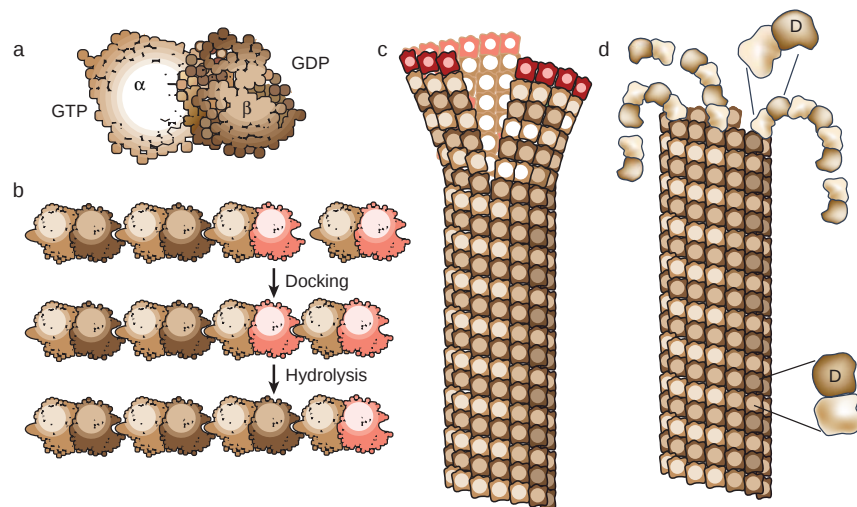
We can now summarize with some confidence the relationship between GTP hydrolysis and the structural changes at the end of the microtubule. First, GTP-tubulin polymerizes onto the end of the microtubule (Fig. 2a). Second, docking of the α -subunit with the β -subunit of the lattice-attached dimer completes the hydrolysis pocket, triggering GTP hydrolysis (Fig. 4b). Third, GTP hydrolysis induces a bend within the subunit (or between subunits), inducing curvature in the lattice and destabilizing the microtubule (Fig. 4c). Thus the bending of the subunit induced by GTP hydrolysis is analogous to the powerstroke of a motor — the fuel driving the polymerization engine is GTP-tubulin binding to the end of the microtubule, whereas the fuel driving the depolymerization engine is release of mechanical strain from the lattice.

Proteins that bind to microtubule ends

Coupling of dynamic microtubule ends to cellular structures requires proteins with unusual properties. If a protein binds to the end of a shrinking microtubule, will it not detach as the tubulin dimers at the end detach? Conversely, if a protein binds to the end of a growing microtubule, will it not block the association of additional tubulin dimers?

Proteins that modulate microtubule dynamics have been known traditionally as microtubule-associated proteins or MAPs⁴⁷. Such proteins, originally isolated from bovine brain, but since identified in all systems studied, increase the growth rate and prevent microtubule catastrophes. So far, studies of MAPs have told us little about the mechanisms by which proteins modulate the dynamics of the microtubule ends. The reason is that they bind all along the microtubule lattice, yet we expect that their effect on dynamics should take place only at the microtubule end. A significant step forward in understanding the dynamics of the plus end was taken with the introduction of green fluorescent protein (GFP) technology to describe proteins that specifically target microtubule ends and in many cases mediate their dynamics^{48–50}. Two distinct classes of end-binding proteins have been described: the MCAKs (for mitotic centromere-associated kinesins),

Figure 4 Model for how the GTP hydrolysis cycle is coupled to structural changes in the microtubule. **a**, Atomic structure of the tubulin dimer as seen in the wall of the protofilament. **b**, Docking of the α - β subunit to the microtubule end. Residues from the incoming α -subunit trigger hydrolysis of the GTP bound to the lattice-attached β -subunit. **c**, **d**, Microtubules at growing ends contain sheets of protofilaments while microtubules at shrinking ends curl. The straight–bent transition is also shown in panel **d**. The GTP dimer is thought to have a straight conformation that fits nicely into the straight wall of the microtubule. Hydrolysis of GTP induces a bend in the subunit, but this bend is constrained within the lattice. The constraint places stress on the lattice, which is released during depolymerization, allowing the protofilament to adopt a curled conformation.



which bind to microtubule ends and destabilize them (Fig. 5a), and the plus-end-binding proteins (or +TIPs⁴⁸), which bind to the growing end of the microtubule and at least in some cases stabilize the microtubule during its growth phase (Fig. 5c).

MCAK/Kin I kinesins

The best understood end-binding proteins are the MCAKs, also called Kin I kinesins. These unusual kinesins^{51,52}, rather than moving along the surface of microtubules like other motor proteins, use energy from ATP hydrolysis to bind to the ends of microtubules, remove tubulin subunits and thus trigger depolymerization^{53,54}. Removal of the *Xenopus* MCAK (XKCM1) from egg extracts dramatically increases the size of the microtubule arrays⁵⁵ by suppressing catastrophes⁵⁶. Overexpressing MCAK in tissue culture cells leads to an almost complete loss of microtubules⁵⁷, perhaps by increasing catastrophes. The localization of MCAK at kinetochores suggests that they could trigger depolymerization during mitosis⁵⁸. It has recently been shown that the combination of XKCM1 and a MAP (XMAP215) can reconstitute the physiological properties of dynamic instability *in vitro*⁵⁹. Thus it seems that, by increasing the catastrophe rate, MCAKs are central to the generation of dynamic microtubules inside cells.

How might the interaction of MCAKs with the end of a growing microtubule convert it to a shrinking one? In the presence of non-hydrolysable ATP analogues, MCAK-family proteins bind to the ends of microtubules and form curled protofilaments — the rams' horns^{53,60,61}. These observations suggest that MCAK proteins bind preferentially to the bent form of the tubulin dimer (Fig. 5b). Even growing microtubules are expected to have a small flair at their ends, owing to internal strain of the GTP subunits⁶², and MCAK may discriminate between the ends of a microtubule and the lattice (that is, the lateral surface) by recognizing these slightly bent subunits in the flared region. A plausible hypothesis for how MCAK destabilizes a growing microtubule is that, after it binds to the end, it causes additional bending, inducing the formation of the curl, which weakens the association of the terminal GTP–tubulin dimer and catalyses its dissociation into solution. Thus by triggering release of GTP subunits from the end of the microtubule, MCAK gates the release of the strained GDP subunits that were trapped in the lattice.

Plus-end-binding proteins

The first bona fide plus-end-binding protein described was CLIP-170, a linker between membranes and microtubules⁶³. As

microtubules grow in the presence of GFP–CLIP-170, bright patches can be seen at the growing end; these patches then disappear when the microtubule stops growing^{63,64} (Fig. 5c). Both the *S. pombe*⁶⁵ and the *S. cerevisiae*⁶⁶ homologues of CLIP-170 have also been shown to target microtubule ends. Work in tissue culture cells illustrates the interaction between CLIP-170 and dynamic microtubules. Here, microtubules growing from centrosomes initially exhibit similar dynamic instability properties as described *in vitro*⁶⁷. That is, they have a low catastrophe rate and if a microtubule does catastrophe, it usually shrinks back to the nucleation centre because the rescue rate is also low. But when a microtubule reaches the cell periphery, the stability of its plus end changes markedly. Here, microtubules that undergo catastrophe rapidly rescue, and microtubules close to the membrane show frequent fluctuations between phases of growing and shrinking⁶⁷. This is thought to allow the microtubules to adapt rapidly to changes in cell shape. Recent work has suggested that these rescue events near the cell periphery are determined by CLIP-170. Removal of CLIP-170 binding to microtubules by dominant negative constructs inhibits rescue of microtubules near the cortex, thus preventing the formation of stable populations of microtubules⁶⁴.

In *S. pombe*, removal of CLIP-170 leads to an increase in catastrophe rates so that few microtubules reach the end of the cell⁶⁵. As a result, polarized growth that takes place at the end of the cell is impaired, leading to an aberrant cell morphology. The results in yeast suggest that microtubule dynamics play a role in cell signalling by providing a mechanism for the targeting of signals (perhaps by association with the CLIP-170 complex) that are necessary for polarized growth. Studies on the interaction between microtubules and focal contacts provide further evidence for a role of the microtubule end in cell signaling⁶⁸.

Since the discovery of CLIP-170, many more plus-end-binding proteins have been identified^{48,69,70}. CLASP proteins target microtubule ends by binding to CLIP-170 (ref. 71). EB1 has been shown to bind to the tips of growing microtubules⁴⁹, where it stabilizes the polymer in mitosis by preventing catastrophes⁷² and may recruit adenomatous polyposis coli (APC) to the microtubule end⁴⁹. Stu2, the XMAP215 homologue in *S. cerevisiae*, also targets the ends of growing microtubules⁷³.

The discovery of these different end-binding proteins is beginning to shed light on how microtubule ends can couple to the cortex and thus mediate mechanical work. In *S. cerevisiae*, the Kar9 protein, which may be the yeast analogue of APC, links microtubule ends to the cortex. The binding of Kar9 to microtubule ends is dependent on the

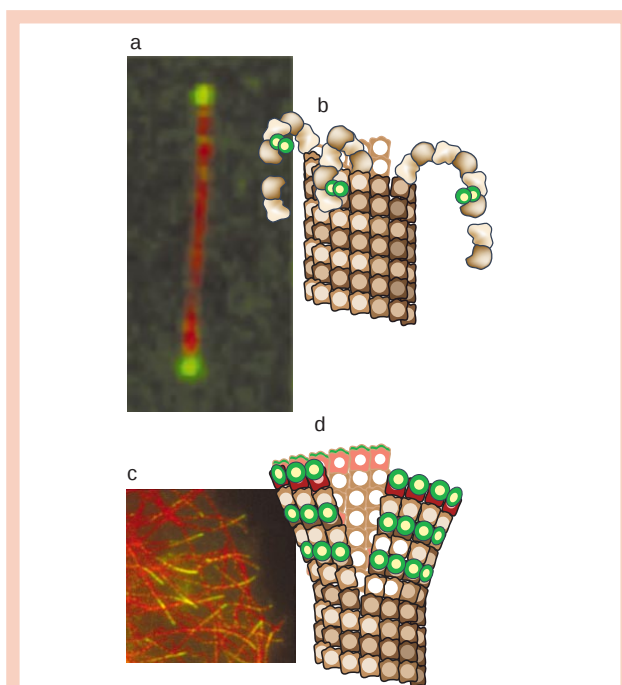


Figure 5 Proteins that recognize microtubule ends. **a**, GFP–MCAK bound to microtubule ends *in vitro*. **b**, Model for MCAK (green) binding to the lattice. **c**, GFP–CLIP-170 bound to the ends of growing microtubules in cells. The yellow segments represent GFP–CLIP-170 at microtubule ends, and the red is microtubules. (Image courtesy of A. Akhmanova.) **d**, Model for CLIP-170 (green) binding to microtubule ends.

end-binding protein EB1. Thus EB1 loads Kar9 onto microtubule ends. When these Kar9 ends reach the cell periphery, they apparently interact with the cortex via cytoplasmic myosin^{23,25,74,75}. This interaction provides a secure coupling so that depolymerization at the plus end pulls the spindle pole body towards the bud. It has been suspected for some time that microtubules also interact with the dynein/dynactin complex at the cortex⁷⁶. Recent work suggests that the dynein/dynactin complex associates with CLIP-170 and in this way targets microtubule ends⁷⁷. Because the dynein/dynactin complex can bind to the actin cortex, this may provide the molecular linkage that allows the complex to mediate spindle positioning in various species^{21,76}.

Plus-end-binding proteins bind to microtubule ends in a different manner to MCAK. The original studies with CLIP-170 suggested a mechanism by which CLIP-170 loads on with the tubulin dimer, but the observation of sheets at the ends of growing microtubules (Fig. 4c) suggests another possible mechanism. Examination of the dynamics of CLIP-170 plus-end segments shows them to be about 1 μm long⁶³. Sheets of over 1 μm in length have been measured in *Xenopus* egg extracts⁴¹. An attractive possibility is that CLIP-170-like proteins target the sheets of microtubules and dissociate as the sheet closes into a tube (Fig. 5d). Recent studies with EB1 provide additional support for this idea⁷², as small sheet-like structures can be seen at the ends of microtubules in the presence of GFP–EB1. A unifying hypothesis could be that the end-binding proteins act by binding to and stabilizing the appropriate end structure — the curled protofilament in the case of MCAK and the sheet in the case of CLIP-170. The sheet stabilizes the end against depolymerization whereas the curl destabilizes the microtubule end.

Outlook

It is clear that studies on the relationship between the biochemistry of end-binding proteins and the physiology of the microtubule end are at an early stage. Do the proteins modulate the structure of the end?

Do they change the rate of GTP hydrolysis? Do they catalyse nucleotide exchange? Do they induce structural transitions as suggested by the work with MCAKs? All these mechanisms are possible and it will be crucial to reconstitute the activities of these proteins with dynamic microtubules, as has been done for the proteins that regulate the dynamics of the actin cytoskeleton⁷⁸. The recent reconstitution of microtubule dynamics using a three-component system of tubulin, MCAK and XMAP215 is a step in this direction⁵⁹. □

doi:10.1038/nature01600

- Doxsey, S. Re-evaluating centrosome function. *Nature Rev. Mol. Cell Biol.* **2**, 688–698 (2001).
- Vinh, D. B. N., Kern, J. W., Hancock, W. O., Howard, J. & Davis, T. N. Reconstitution and characterization of budding yeast γ -tubulin complex. *Mol. Biol. Cell* **13**, 1144–1157 (2002).
- Meads, T. & Schroer, T. A. Polarity and nucleation of microtubules in polarized epithelial cells. *Cell Motil. Cytoskel.* **32**, 273–288 (1995).
- Tassin, A. & Bornens, M. Centrosome structure and microtubule nucleation in animal cells. *Biol. Cell* **91**, 343–354 (1999).
- Mitchison, T. & Kirschner, M. Dynamic instability of microtubule growth. *Nature* **312**, 237–242 (1984).
- Cassimeris, L., Pryer, N. K. & Salmon, E. D. Real-time observations of microtubule dynamic instability in living cells. *J. Cell Biol.* **107**, 2223–2231 (1988).
- Sammak, P. J. & Borisy, G. G. Direct observation of microtubule dynamics in living cells. *Nature* **332**, 724–726 (1988).
- Belmont, L. D., Hyman, A. A., Sawin, K. E. & Mitchison, T. J. Real-time visualization of cell cycle-dependent changes in microtubule dynamics in cytoplasmic extracts. *Cell* **62**, 579–589 (1990).
- Kinoshita, K., Habermann, B. & Hyman, A. A. XMAP215: a key component of the dynamic microtubule cytoskeleton. *Trends Cell Biol.* **12**, 267–273 (2002).
- Kirschner, M. & Mitchison, T. Beyond self-assembly: from microtubules to morphogenesis. *Cell* **45**, 329–342 (1986).
- Karsenti, E. Mitotic spindle morphogenesis in animal cells. *Semin. Cell Biol.* **2**, 251–260 (1991).
- Hyman, A. A. & Karsenti, E. Morphogenetic properties of microtubules and mitotic spindle assembly. *Cell* **84**, 401–410 (1996).
- Hayden, J. H., Bowser, S. S. & Rieder, C. L. Kinetochore capture astral microtubules during chromosome attachment to the mitotic spindle: direct visualization in live newt lung cells. *J. Cell Biol.* **111**, 1039–1045 (1990).
- Karsenti, E. & Vernos, I. The mitotic spindle: a self-made machine. *Science* **294**, 543–547 (2001).
- Hirokawa, N. Kinesin and dynein superfamily proteins and the mechanism of organelle transport. *Science* **279**, 519–526 (1998).
- Howard, J. *Mechanics of Motor Proteins and the Cytoskeleton* (Sinauer Associates, Sunderland, MA, 2001).
- Vale, R. D. & Milligan, R. A. The way things move: looking under the hood of molecular motor proteins. *Science* **288**, 88–95 (2000).
- Rieder, C. L. & Salmon, E. D. The vertebrate cell kinetochore and its roles during mitosis. *Trends Cell Biol.* **8**, 310–318 (1998).
- Inoue, S. & Salmon, E. D. Force generation by microtubule assembly/disassembly in mitosis and related movements. *Mol. Biol. Cell* **6**, 1619–1640 (1995).
- Doe, C. Q. & Bowerman, B. Asymmetric cell division: fly neuroblast meets worm zygote. *Curr. Opin. Cell Biol.* **13**, 68–75 (2001).
- Gonczy, P. Mechanisms of spindle positioning: focus on flies and worms. *Trends Cell Biol.* **12**, 332–339 (2002).
- Tran, P. T., Marsh, L., Doye, V., Inoue, S. & Chang, F. A mechanism for nuclear positioning in fission yeast based on microtubule pushing. *J. Cell Biol.* **153**, 397–412 (2001).
- Kusch, J., Meyer, A., Snyder, M. P. & Barral, Y. Microtubule capture by the cleavage apparatus is required for proper spindle positioning in yeast. *Genes Dev.* **16**, 1627–1639 (2002).
- Yeh, E. *et al.* Dynamic positioning of mitotic spindles in yeast: role of microtubule motors and cortical determinants. *Mol. Biol. Cell* **11**, 3949–3961 (2000).
- Liakopoulos, D., Kusch, J., Grava, S., Vogel, J. & Barral, Y. Asymmetric loading of Kar9 onto spindle poles and microtubules ensures proper spindle alignment. *Cell* **112**, 561–574 (2003).
- Maekawa, H., Usui, T., Knop, M. & Schiebel, E. Yeast Cdk1 translocates to the plus end of cytoplasmic microtubules to regulate bud cortex interactions. *EMBO J.* **22**, 438–449 (2003).
- Fygenson, D. K., Marko, J. F. & Libchaber, A. Mechanics of microtubule-based membrane extension. *Phys. Rev. Lett.* **79**, 4497–4500 (1997).
- Dogterom, M. & Yurke, B. Measurement of the force-velocity relation for growing microtubules. *Science* **278**, 856–860 (1997).
- Coue, M., Lombillo, V. A. & McIntosh, J. R. Microtubule depolymerization promotes particle and chromosome movement *in vitro*. *J. Cell Biol.* **112**, 1165–1175 (1991).
- Faivre-Moskalenko, C. & Dogterom, M. Dynamics of microtubule asters in microfabricated chambers: the role of catastrophes. *Proc. Natl Acad. Sci. USA* **99**, 16788–16793 (2002).
- Holy, T. E., Dogterom, M., Yurke, B. & Leibler, S. Assembly and positioning of microtubule asters in microfabricated chambers. *Proc. Natl Acad. Sci. USA* **94**, 6228–6231 (1997).
- Erickson, H. P. & O'Brien, E. T. Microtubule dynamic instability and GTP hydrolysis. *Annu. Rev. Biophys. Biomol. Struct.* **21**, 145–166 (1992).
- Hyman, A. A., Salser, S., Drechsel, D. N., Unwin, N. & Mitchison, T. J. Role of GTP hydrolysis in microtubule dynamics: information from a slowly hydrolyzable analogue, GMPCPP. *Mol. Biol. Cell* **3**, 1155–1167 (1992).
- Desai, A. & Mitchison, T. J. Microtubule polymerization dynamics. *Annu. Rev. Cell Dev. Biol.* **13**, 83–117 (1997).
- Mitchison, T. J. Localization of an exchangeable GTP binding site at the plus end of microtubules. *Science* **261**, 1044–1047 (1993).
- Nogales, E., Whittaker, M., Milligan, R. A. & Downing, K. H. High-resolution model of the microtubule. *Cell* **96**, 79–88 (1999).
- Drechsel, D. N. & Kirschner, M. W. The minimum GTP cap required to stabilize microtubules. *Curr. Biol.* **4**, 1053–1061 (1994). [Published erratum appears in *Curr. Biol.* **5**, 215 (1995).]

38. Caplow, M. & Shanks, J. Evidence that a single monolayer tubulin-GTP cap is both necessary and sufficient to stabilize microtubules. *Mol. Biol. Cell* **7**, 663–675 (1996).
39. Geeves, M. A. & Holmes, K. C. Structural mechanism of muscle contraction. *Annu. Rev. Biochem.* **68**, 687–728 (1999).
40. Chretien, D., Fuller, S. D. & Karsenti, E. Structure of growing microtubule ends: two-dimensional sheets close into tubes at variable rates. *J. Cell Biol.* **129**, 1311–1328 (1995).
41. Arnal, I., Karsenti, E. & Hyman, A. A. Structural transitions at microtubule ends correlate with their dynamic properties in *Xenopus* egg extracts. *J. Cell Biol.* **149**, 767–774 (2000).
42. Mandelkow, E. M., Mandelkow, E. & Milligan, R. A. Microtubule dynamics and microtubule caps: a time-resolved cryo-electron microscopy study. *J. Cell Biol.* **114**, 977–991 (1991).
43. Melki, R., Carlier, M. F., Pantaloni, D. & Timasheff, S. N. Cold depolymerization of microtubules to double rings: geometric stabilization of assemblies. *Biochemistry* **28**, 9143–9152 (1989).
44. Hyman, A. A., Chretien, D., Arnal, I. & Wade, R. H. Structural changes accompanying GTP hydrolysis in microtubules: information from a slowly hydrolyzable analogue guanylyl-(α,β)-methylene-diphosphonate. *J. Cell Biol.* **128**, 117–125 (1995).
45. Muller-Reichert, T., Chretien, D., Severin, F. & Hyman, A. A. Structural changes at microtubule ends accompanying GTP hydrolysis: information from a slowly hydrolyzable analogue of GTP, guanylyl-(α,β)-methylenediphosphonate. *Proc. Natl Acad. Sci. USA* **95**, 3661–3666 (1998).
46. Gigant, B. *et al.* The 4 Å X-ray structure of a tubulin:stathmin-like domain complex. *Cell* **102**, 809–816 (2000).
47. Andersen, S. S. Spindle assembly and the art of regulating microtubule dynamics by MAPs and Stathmin/Op18. *Trends Cell Biol.* **10**, 261–267 (2000).
48. Schuyler, S. C. & Pellman, D. Microtubule “plus-end-tracking proteins”: The end is just the beginning. *Cell* **105**, 421–424 (2001).
49. Mimori-Kiyosue, Y. & Tsukita, S. Where is APC going? *J. Cell Biol.* **154**, 1105–1109 (2001).
50. Tirnauer, J. S. & Bierer, B. E. EB1 proteins regulate microtubule dynamics, cell polarity, and chromosome stability. *J. Cell Biol.* **149**, 761–766 (2000).
51. Kim, A. J. & Endow, S. A. A kinesin family tree. *J. Cell Sci.* **113**, 3681–3682 (2000).
52. Lawrence, C. J., Malmberg, R. L., Muszynski, M. G. & Dawe, R. K. Maximum likelihood methods reveal conservation of function among closely related kinesin families. *J. Mol. Evol.* **54**, 42–53 (2002).
53. Desai, A., Verma, S., Mitchison, T. J. & Walczak, C. E. Kin I kinesins are microtubule-destabilizing enzymes. *Cell* **96**, 69–78 (1999).
54. Hunter, A. W. *et al.* The kinesin-related protein MCAK is a microtubule depolymerase that forms an ATP-hydrolyzing complex at microtubule ends. *Mol. Cell* (in the press).
55. Walczak, C. E., Mitchison, T. J. & Desai, A. XKCM1: a *Xenopus* kinesin-related protein that regulates microtubule dynamics during mitotic spindle assembly. *Cell* **84**, 37–47 (1996).
56. Tournibize, R. *et al.* Control of microtubule dynamics by the antagonistic activities of XMAP215 and XKCM1 in *Xenopus* egg extracts. *Nature Cell Biol.* **2**, 13–19 (2000).
57. Maney, T., Wagenbach, M. & Wordeman, L. Molecular dissection of the microtubule depolymerizing activity of mitotic centromere-associated kinesin. *J. Biol. Chem.* **276**, 34753–34758 (2001).
58. Maney, T., Hunter, A. W., Wagenbach, M. & Wordeman, L. Mitotic centromere-associated kinesin is important for anaphase chromosome segregation. *J. Cell Biol.* **142**, 787–801 (1998).
59. Kinoshita, K., Arnal, I., Desai, A., Drechsel, D. N. & Hyman, A. A. Reconstitution of physiological microtubule dynamics using purified components. *Science* **294**, 1340–1343 (2001).
60. Moores, C. A. *et al.* A mechanism for microtubule depolymerization by KinI kinesins. *Mol. Cell* **9**, 903–909 (2002).
61. Niederstrasser, H., Salehi-Had, H., Gan, E. C., Walczak, C. & Nogales, E. XKCM1 acts on a single protofilament and requires the C terminus of tubulin. *J. Mol. Biol.* **316**, 817–828 (2002).
62. Janosi, I. M., Chretien, D. & Flyvbjerg, H. Structural microtubule cap: stability, catastrophe, rescue, and third state. *Biophys. J.* **83**, 1317–1330 (2002).
63. Perez, F., Diamantopoulos, G. S., Stalder, R. & Kreis, T. E. CLIP-170 highlights growing microtubule ends in vivo. *Cell* **96**, 517–527 (1999).
64. Komarova, Y. A., Akhmanova, A. S., Kojima, S.-i., Galjart, N. & Borisy, G. G. Cytoplasmic linker proteins promote microtubule rescue in vivo. *J. Cell Biol.* **159**, 589–599 (2002).
65. Brunner, D. & Nurse, P. CLIP170-like tip1p spatially organizes microtubular dynamics in fission yeast. *Cell* **102**, 695–704 (2000).
66. Lin, H. *et al.* Polyploids require Bik1 for kinetochore-microtubule attachment. *J. Cell Biol.* **155**, 1173–1184 (2001).
67. Komarova, Y. A., Vorobjev, I. A. & Borisy, G. G. Life cycle of MTs: persistent growth in the cell interior, asymmetric transition frequencies and effects of the cell boundary. *J. Cell Sci.* **115**, 3527–3539 (2002).
68. Small, J. V. & Kaverina, I. Microtubules meet substrate adhesions to arrange cell polarity. *Curr. Opin. Cell Biol.* **15**, 40–47 (2003).
69. Sawin, K. E. Microtubule dynamics: the view from the tip. *Curr. Biol.* **10**, R860–R862 (2000).
70. Schroer, T. A. Microtubules don and doff their caps: dynamic attachments at plus and minus ends. *Curr. Opin. Cell Biol.* **13**, 92–96 (2001).
71. Akhmanova, A. *et al.* Clasps are CLIP-115 and -170 associating proteins involved in the regional regulation of microtubule dynamics in motile fibroblasts. *Cell* **104**, 923–935 (2001).
72. Tirnauer, J. S., Grego, S., Salmon, E. D. & Mitchison, T. J. EB1-microtubule interactions in *Xenopus* egg extracts: role of EB1 in microtubule stabilization and mechanisms of targeting to microtubules. *Mol. Biol. Cell* **13**, 3614–3626 (2002).
73. He, X., Rines, D. R., Espelin, C. W. & Sorger, P. K. Molecular analysis of kinetochore-microtubule attachment in budding yeast. *Cell* **106**, 195–206 (2001).
74. Lee, L. *et al.* Positioning of the mitotic spindle by a cortical-microtubule capture mechanism. *Science* **287**, 2260–2262 (2000).
75. Maekawa, H., Usui, T., Knop, M. & Schiebel, E. Yeast Cdk1 translocates to the plus end of cytoplasmic microtubules to regulate bud cortex interactions. *EMBO J.* **22**, 438–449 (2003).
76. Hildebrandt, E. R. & Hoyt, M. A. Mitotic motors in *Saccharomyces cerevisiae*. *Biochim. Biophys. Acta* **1496**, 99–116 (2000).
77. Valetti, C. *et al.* Role of dynactin in endocytic traffic: effects of dynactin overexpression and colocalization with CLIP-170. *Mol. Biol. Cell* **10**, 4107–4120 (1999).
78. Loisel, T. P., Boujemaa, R., Pantaloni, D. & Carlier, M. F. Reconstitution of actin-based motility of *Listeria* and *Shigella* using pure proteins. *Nature* **401**, 613–616 (1999).

Acknowledgements We thank Y. Barral and G. Borisy for sharing unpublished data. Research in the authors' laboratories is supported by the Human Frontier Science Program, the National Institutes of Health and the Max Planck Gesellschaft.

Molecular motors

Manfred Schliwa & Günther Woehlke

Adolf Butenandt Institut, Zellbiologie, Universität München, Schillerstrasse 42, 80336 München, Germany
(e-mail: schliwa@bio.med.uni-muenchen.de)

Life implies movement. Most forms of movement in the living world are powered by tiny protein machines known as molecular motors. Among the best known are motors that use sophisticated intramolecular amplification mechanisms to take nanometre steps along protein tracks in the cytoplasm. These motors transport a wide variety of cargo, power cell locomotion, drive cell division and, when combined in large ensembles, allow organisms to move. Motor defects can lead to severe diseases or may even be lethal. Basic principles of motor design and mechanism have now been derived, and an understanding of their complex cellular roles is emerging.

Molecular motors are amazing biological machines that are responsible for most forms of movement we encounter in the cellular world. Three types of cytoplasmic motors are known: myosins, which move on actin filaments, and dyneins and kinesins, which use microtubules as tracks. The mechanism they use to convert chemical energy into mechanical work is both simple and ingenious. In all three motor classes, ATP hydrolysis causes a small conformational change in a globular motor domain that is amplified and translated into movement with the aid of accessory structural motifs. Additional domains outside the motor unit are responsible for dimerization, regulation and interactions with other molecules (Fig. 1).

This modular design of motors has given rise to considerable complexity so that each of the three motors comprises a superfamily whose members may vary appreciably in makeup and function. Today, we can distinguish at least 18 different classes of myosins, 10 different families of kinesins, and 2 groups of dyneins, each with up to several dozen members. The complement of motors varies widely between different organisms. Yeast, for example, gets by with 6 kinesins, 5 myosins and 1 dynein, whereas mammals have genes for over 40 kinesins, 40 myosins and more than a dozen dyneins. These numbers may easily be tripled as a result of post-translational modifications or varied combinations of associated proteins. Many motors are not yet characterized, and clear functions are assigned to only a small subset. Nevertheless, remarkable insights into motor mechanochemistry and function have been gained. This introductory overview highlights recent developments; for a compilation of comprehensive reviews, see ref. 1.

Motor mechanochemistry

Conformational changes

Our understanding of the molecular mechanisms that convert chemical energy into movement is most advanced for representatives of the myosin and kinesin families. High-resolution crystal structures of the motor domain uncovered an unexpected relationship between these two classes of motors: the region surrounding the ATP-binding pocket is virtually identical in structure, although sequence homology is restricted to only a few key residues. The architecture of the active site further revealed a relationship to the G proteins, suggesting that these three classes of molecules are of common evolutionary origin². This notion recently received support from molecular dynamics simulations

suggesting that G proteins — usually mediators in signalling pathways — may be able to generate force³.

Among the various families of kinesins and myosins we find motors that work as monomers, dimers, trimers or tetramers, move to the plus end or the minus end of their track, and take just one or many steps before dissociating. Despite this wide spectrum of behaviours, in all motors the initial events in the generation of movement are similar and can be explained by stepwise amplification (Fig. 2a, b).

The primary event, the loss of the γ -phosphate group from ATP, leaves a space of approximately 0.5 nm, which is thought to cause a rearrangement of conserved structural elements flanking the ATP-binding site. This rearrangement, which represents the first level of amplification, is coordinated with structural changes in the track-binding site. Interruption of this coordination uncouples ATP hydrolysis from track binding^{4,5}. The next level of amplification involves communication of the conformational change in the active site to carboxy-terminal structural components that may be viewed as mechanical amplifiers. Here myosins and kinesins differ. In many myosins, the mechanical amplifier is an α -helix of variable length stabilized by light chains. Based on crystal structures in different nucleotide states, this rigid structure acts as a lever that swings through an angle of up to 70° (refs 6, 7). The lever swing is believed to be the ultimate cause for the working stroke⁸. Accordingly, motors with longer necks take larger steps and move faster^{9,10}. In conventional kinesins, the amplifier is a short, flexible stretch of ~10 amino acids that can be either docked to the motor core or flexible and free¹¹. The mobility of this neck linker, possibly coupled to a rotation of parts of the motor domain¹², is believed to drive kinesin movement. Thus the structural features that sense and transmit hydrolysis-dependent changes are similar in the two motors, but translation into a large-scale conformational change apparently involves rotation of a rigid stalk in myosin and repositioning of a flexible element in kinesin¹³.

Mechanistic analysis of the dynein motor is severely hampered by the lack of a high-resolution structure. It is clear though that, based on sequence features, the molecular design of dyneins is fundamentally different from myosins and kinesins. The motor domain of dynein comprises a ring of six AAA-ATPase modules, members of a widespread and highly diverse superfamily of proteins. ATP-dependent conformational changes in the ring of AAA-modules are believed to be transmitted to a stalk that carries the microtubule-binding site at its tip¹⁴. A swing in the position of this stalk leads to a ~15-nm displacement of the tip (Fig. 2c)¹⁵.

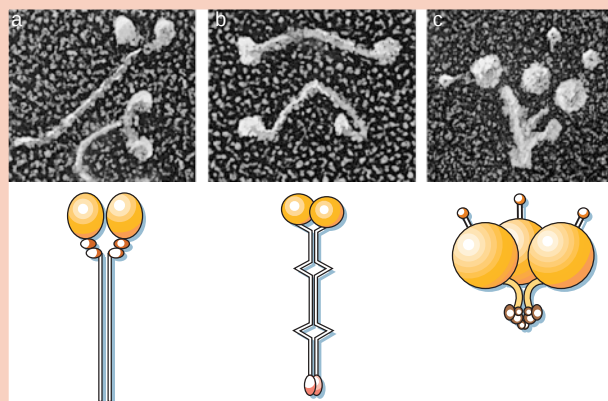


Figure 1 Representative cytoskeletal motors. **a**, Myosin II; **b**, conventional kinesin; **c**, ciliary dynein. The top row shows high-resolution electron micrographs of quick-frozen, rotary-shadowed individual molecules (images courtesy of J. Heuser). Corresponding schematic overviews are shown below. Motor domains are in yellow, associated proteins are shown in brown, and coiled-coil domains are represented by parallel black lines. For detailed overviews of the superfamilies of myosin and kinesin motors, see the myosin home page (<http://www.mrc-lmb.cam.ac.uk/myosin/myosin.html>) and the kinesin home page (<http://www.proweb.org/kinesin/>).

Although superficially resembling a swinging lever-arm movement, the structural and molecular basis of this force-generating ‘power stroke’ differs markedly from the conformational changes in myosins and kinesins.

Stepping

The conversion of these conformational changes into a step (or series of steps) leads us to the next level of complexity. Two fundamentally different behaviours of motors can be distinguished. In one, a single motor molecule can move along the track for long distances without detaching, a behaviour referred to as processivity. In the second, motors lose contact to the track usually after one cycle and therefore are non-processive. These modes of operation are physiological adaptations to different cellular functions. Processive motors are individualists, whereas non-processive motors often work as a team; the former hold on to the track for as long as possible, whereas the latter are optimized for brief, fast interactions.

Conventional kinesin, perhaps the best example of a strictly processive motor, is a dimer that interlaces the reaction cycles of the two heads. One head is tightly bound to the microtubule for at least half of the time of an ATPase cycle, and the two heads are kept out of phase¹⁶. Most models predict a ‘hand-over-hand’ cycle where the free head moves towards a new binding site past the bound head, consuming one ATP per step. An alternative model proposes an ‘inchworm’ type of asymmetrical stepping with a ‘front’ and a ‘back’ head¹⁷. In either model, a phase must exist where both heads are bound to the microtubule. Because a crystal structure of dimeric kinesin places the two motor domains in an unfavourable orientation only 5 nm apart¹⁸, major rearrangements of adjacent domains are required during stepping. As discussed above, these rearrangements may be accommodated by the flexible neck-linker domain¹¹. Partial unravelling of the coiled-coil neck may also be involved in some¹⁹, but not all²⁰, kinesins.

The paradigm for a non-processive motor is muscle myosin II, which uses a lever arm to generate its working stroke. As conventional kinesin it is dimeric, but unlike conventional kinesin, the two heads do not cooperate, and the interaction with the track takes up less than one-tenth of the time of an ATPase cycle. Both factors contribute to its non-processivity. This mode of operation of myosin II makes sense because, in the sarcomeric ensemble, motors that remain bound to actin after their power stroke would slow down the entire system.

There are, however, myosins that possess a rigid lever arm (like myosin II) and combine it with head–head coordination (like kinesin) to operate processively. The best example is myosin V. Its six light-chain-binding sites in the neck create an extraordinarily long lever arm that enables a large step. Indeed, myosin V’s step size is ~36 nm, which corresponds to the pitch of the actin helix²¹, and its velocity and ATPase activity are consistent with the hydrolysis of one molecule of ATP per 36-nm step²². This large stride apparently requires contributions from two different mechanisms: a working stroke of only ~25 nm, and thermally driven diffusion, which contributes the missing 11 nm²³. A similar ‘composite’ mechanism also seems to operate in other motors such as myosin VI where the short power stroke serves primarily to bias the reverse directionality of this motor while thermal motion drives its movement²⁴. In both motors, the activities of the two heads must be strictly coordinated, which may be achieved via elastic strain exerted on the rear head by the curved neck of the forward head when both heads are bound²⁵.

Processive movement was generally believed to require dimeric motors. It therefore came as a surprise when monomeric KIF1A kinesin²⁶, monomeric class IXb myosin²⁷ and monomeric inner arm dynein²⁸ were suggested to move processively. But their mode of processivity differs from that of dimeric motors. For example, *in vitro*, monomeric KIF1A diffuses back and forth for several seconds when bound to microtubules, with a net movement towards the microtubule plus end. The key to this behaviour is the presence of a positively charged loop that interacts with the negatively charged C terminus of tubulin. This loop acts as a tether while the power stroke of KIF1A provides the push that biases diffusion towards the microtubule plus end²⁹. Performance-enhancing charge interactions may also help to keep dimeric motors ‘on track’³⁰. Whether charge-dependent tethering is the key to understanding monomer processivity remains in doubt, as other members of the KIF1 family that also possess the ominous K-loop are non-processive³¹. Moreover, KIF1A-like kinesins may actually dimerize under *in vivo* conditions³², relegating the mode of monomer movement to a mechanistically intriguing, but physiologically irrelevant, *in vitro* phenomenon.

A general conclusion emerging from studies on processive motors is that moving along the track may entail both a mechanical component and a diffusive component, with different motors using different proportions of each. Some motors rely largely on rigid conformational changes and tight coupling, with a relatively small contribution from diffusional searching. Others seem to have a relatively large contribution from diffusion, which alters their manner and form of processivity. In both, the diffusional component is supported by secondary ‘tethering’ sites that enhance motor performance.

Directionality

Most cell biologists would have been rather comfortable with the idea that a given superfamily of motors moves in one direction only. This comforting thought was shattered with the discovery of minus-end-directed kinesin-like proteins and a minus-end-directed myosin, leaving dynein as the last hope for a unidirectional motor superfamily.

All minus-end-directed kinesins studied so far have the motor domain at the C terminus, as opposed to the N terminus in plus-end motors. The two heads of ncd, for example, are tightly associated with the neck coiled-coil³³, which alters head–neck interaction, a key factor in determining directionality. When motor domains of forward and reverse motors are swapped, the resulting chimaeras adopt the direction of movement specified by the neck³⁴. Movement of the chimaeras is usually slow and points to an intrinsic but weak plus-end bias even in a minus-end motor. Convincing evidence for the importance of the neck region in directional determination came from the analysis of a point mutant in the ncd neck that completely lacks directionality, switching stochastically between plus-end and minus-end movement³⁵.

The reversed polarity of the minus-end-directed myosin VI motor was attributed to a unique insertion of 53 amino acids in the converter

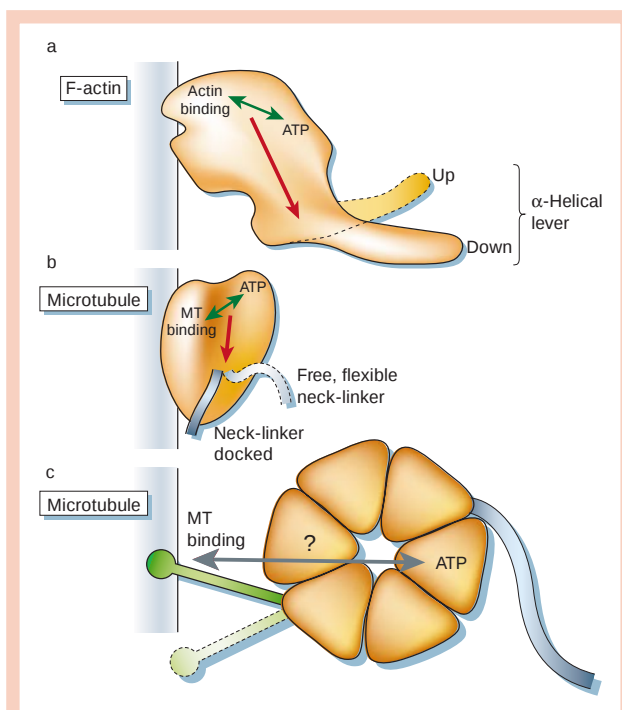


Figure 2 Schematic rendition of the intramolecular communication within one motor domain each of myosin, kinesin and dynein, and translation into a conformational change that leads to movement. In both myosin (**a**) and kinesin (**b**), ATP hydrolysis causes a conformational change to structural elements near the ATP-binding site that is communicated to the track-binding site (green arrow). The information is then relayed (red arrow) via homologous structural elements to a mechanical amplifier. **a**, In myosin the amplifier is a helix stabilized by light chains (not shown) that acts as a swinging lever. **b**, In kinesin the amplifier is a flexible element, the neck linker, that connects the motor domain with the neck helix. This element apparently undergoes a major positional shift, but its precise orientation remains to be determined. **c**, The pathways of intramolecular communication within the dynein motor domain are unknown at present, but the information on ATP hydrolysis is transmitted from one end of the molecule to the stalk that carries the microtubule binding site. The final step apparently involves an angular swing of the stalk.

domain, which is proposed to reverse the direction of the lever-arm swing³⁶. This attractive hypothesis was cast in doubt when results from an analysis of several chimaeras between the opposite-polarity motors myosin V and myosin VI suggested that this insertion is neither necessary nor sufficient for minus-end-directed movement³⁷. So far, studies have failed to show conclusively where the direction-determining regions reside, although it is hoped that clarification will be obtained upon analysis of the crystal structure of the myosin VI motor domain. It seems, however, that the structural basis of directional reversal is fundamentally different in myosin and kinesin motors.

Forces

The concept of serial amplification of structural rearrangements suggests that a minor change of 0.5 nm set off by the presence of absence of a phosphate group can be enlarged up to 36 nm (in myosin V). How big are the forces involved? To measure these forces, ingenious microdevices were developed that operate with unprecedented precision and sensitivity³⁸. Force measurements have been made on only a subset of motors in each superfamily, but they show that the forces developed by kinesin, myosin and dynein motors — about 1–10 pN — are extremely minute by our macroscopic standards. For example, to lift a 5 kg weight, about 10^{13} motors are required. However, in the realm of the cell, these forces are gigantic. A single motor can move an object many times its own size through viscous cytoplasm at near

maximum speed. External forces affect the kinesin cycle, suggesting at least one load-dependent transition, most likely associated with ATP binding³⁹. Improved force-clamp techniques using laser traps equipped with a feedback control⁴⁰ will allow such load-dependent steps to be studied in detail.

Cellular functions

The initial belief that the three types of motors are associated with clearly separate functions (that is, myosin with contraction and movement, dynein with ciliary beating, and kinesin with organelle transport) could not be upheld for long. Now we are aware of, for example, myosins involved in organelle transport, dyneins implicated in vesicle and cell movement, and kinesins required for ciliary function. In addition, we count among their tasks unexpected functions such as signalling, RNA localization and sensory transduction; we are beginning to appreciate their implications in cellular architecture, basic developmental processes and a growing number of diseases; and we know that all three are important in cell division (see review in this issue by Scholey, page 746). This already is an impressive list, but because many motors have not yet been characterized, the full spectrum of cellular roles has yet to be appreciated.

Membrane association and regulation

Members of all three types of cytoskeletal motors are involved in organelle and vesicle transport (for reviews, see ref. 1). To understand these functions, it is essential to determine how motors link up to their cargoes and how transport is regulated. In both processes, non-motor domains and associated proteins have a key role, and a wide spectrum of attachment mechanisms is observed (Fig. 3).

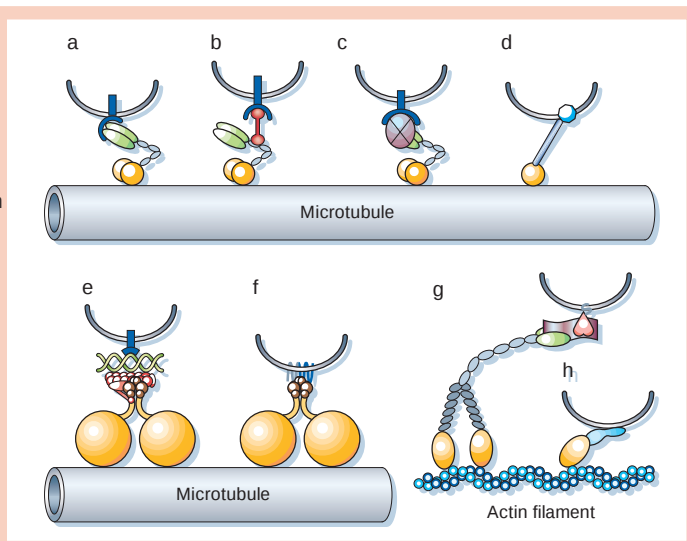
Perhaps the most direct (but seemingly least specific) mechanism of membrane association is linkage to the phospholipid bilayer. Thus, acidic phospholipids are the binding partner for monomeric myosins⁴¹ possessing a basic tail region, whereas a member of the Unc104/KIF1 family of kinesins binds to lipids via a pleckstrin homology domain⁴². This association depends on the presence of phosphatidylinositol 4,5-bisphosphate (PtdIns(4,5)P₂), which promotes clustering of the motor in PtdIns(4,5)P₂-containing rafts. Clustering, in turn, may trigger the onset of transport.

In certain cell types, motors such as conventional kinesin and cytoplasmic dynein can latch onto their cargo via integral membrane proteins. In neurons, the kinesin light chains bind amyloid precursor protein (APP), a transmembrane protein of certain axonally transported vesicles⁴³. This link is of potential medical significance as APP has gained fame as the precursor of a proteolytic fragment that gives rise to amyloid plaques in patients with Alzheimer's disease. Impaired APP transport may well contribute to the development of the disease.

Table 1 Examples of molecular motors involved in disease

Disease or defect	Motor involved	Reference
Myosin myopathies	Myosin II	88
Griscelli syndrome (pigmentation disorder)	Myosin V	89
Hearing loss	Myosin IIIa (ninaC), myosin VI, myosin VIIa, myosin XVa	90
Retinitis pigmentosa (photoreceptor degeneration)	Cytoplasmic dynein, Kinesin Krp85/95	44 91
Primary ciliary dyskinesia	Axonemal dynein	92
Kartagener syndrome (situs inversus)	Axonemal dynein, kinesin Krp85/95	93
Polycystic kidney disease	Dyneins and kinesins	94
Lissencephaly	LIS1 (dynein/dynactin interactor)	95
Charcot-Marie-Tooth disease type 2A	KIF1B (Unc104 kinesin family)	96
Virus transport	Conventional kinesin, cytoplasmic dynein, myosin	97, 98
Anthrax susceptibility	KIF1C (Unc104 kinesin family)	99
Neurodegenerative diseases	Kinesin	100
	Cytoplasmic dynein	101

Figure 3 Types of motor-cargo linkage. **a–d**, kinesin; **e, f**, dynein; **g, h**, myosin. **a**, Interaction between a transmembrane receptor (blue) and kinesin light chains (green)⁴³. **b**, Interaction between a transmembrane receptor and kinesin heavy chains mediated by a linker protein (red)⁴⁷. **c**, Interaction between a transmembrane receptor and kinesin light chains mediated by a linker complex (purple)^{45,46}. **d**, Interaction between membrane phospholipids and a pleckstrin homology domain (blue) in the kinesin-like protein Unc104 (ref. 42). **e**, Interaction between cytoplasmic dynein and an integral membrane protein mediated by the dynactin complex (red) and spectrin (green)^{50,51}. **f**, Direct linkage of the Tctex-1 light chain of dynein with an integral membrane protein, rhodopsin⁴⁴. **g**, Linkage of the tail domain of myosin V to membrane-anchored rab27a (red) via melanophilin (purple)⁴⁸. **h**, Direct interaction of the tail domain of myosin I (blue) with acidic phospholipids⁴¹.



In photoreceptor cells, cytoplasmic dynein, which normally requires the dynactin complex for attachment (see below), binds directly to rhodopsin, an integral membrane protein, with its Tctex-1 light chain⁴⁴. This link, too, is significant as certain rhodopsin mutations inhibit this interaction, leading to retinitis pigmentosa.

The most widespread mode of association with integral membrane proteins occurs via linker proteins, often in the form of large assemblies. Work over the past few years has advanced various attachment modes for all three motor types. For example, conventional kinesin, again via its light chains, interacts with Jun kinase-interacting proteins (JIPs), a class of scaffolding proteins that bind components of the JNK signalling pathway^{45,46}. JIPs, in turn, bind a transmembrane receptor of the low-density lipoprotein receptor family. Certain other kinesin-like proteins likewise use large linker complexes⁴⁷.

Among the myosins, the machinery that links myosin V to cargo is characterized best. In pigment cells, the small GTPase Rab27a and a recently identified Rab-binding protein, melanophilin⁴⁸, attach myosin V to melanosomes. The GTPase binds to membranes first and recruits melanophilin, which then binds myosin V. Melanophilin binding is GTP dependent, thus offering a convenient means of regulating motor–cargo association. This Rab-dependent machinery may well be paradigmatic for myosin–cargo association in other systems. Recent discoveries link Rabs and Rab-like effectors not only to several other myosin motors, but also to kinesins and dynein⁴⁹, thus opening the possibility of a significant functional interdependence of GTPases, motors and membrane traffic.

Finally, a large protein assembly seems to be involved in linking dynein to membranes (Fig. 4). Through its intermediate chains, dynein interacts with a unique activator complex, dynactin, which has the protein p150^{glued} and a short filament of the actin-related protein Arp1 as its most prominent components⁵⁰. Precisely how the dynein–dynactin complex associates with vesicular cargoes is not understood, although in certain circumstances it binds to membrane-associated spectrin⁵¹, making this the most complex linkage machinery known.

These examples, which represent the proverbial tip of the iceberg, indicate a wide spectrum of attachment mechanisms. Direct association with lipids or transmembrane proteins, linkage via an adaptor, or association mediated by complex protein assemblies all have been found. Given that one motor can interact with several different cargoes, there may well be dozens of specific membrane attachment mechanisms matching the dozens of potential cargoes in a cell.

An important issue arising from studies on cargo association is the question of motor regulation in cellular transport. In principle, motor activity can be regulated at two levels: by turning the motor on

or off, and by inhibiting or promoting its association with cargo. Although this is largely uncharted territory, both mechanisms have been encountered in cells. In both, phosphorylation has a significant role, but novel means of regulation exist as well.

Phosphorylation may emerge as a negative regulator of cargo binding of several motors. For example, docking of the globular tail of myosin V onto melanosomes is inhibited by phosphorylation⁵², thus holding up melanosome movement, while phosphorylation of the light intermediate chains releases dynein from membranes⁵³. There are hints that phosphorylation can also regulate kinesin-based organelle movement⁵⁴, but it is unclear whether cargo binding is affected directly.

A radically different mechanism used by kinesin to avert non-productive movement without cargo involves intramolecular folding where the tail inhibits the motor domain. Binding to cargo de-represses tail inhibition and allows the motor to unfold, a process that is critically dependent on a flexible domain in the stalk^{55,56}. This is an attractive mechanism because it couples motor regulation and cargo binding. How the tail inhibits the motor is not known, but a tail motif conserved in all conventional kinesins is crucial^{57,58}. Thus, intramolecular interactions and phosphorylation may complement each other in the regulation of cargo transport.

Coordinating motors

Research on organelle transport took a completely unexpected twist with the demonstration that some organelles can switch tracks and move on either microtubules or actin filaments⁵⁹. In amphibian melanophores, for example, heterotrimeric kinesin and myosin V cooperate in the dispersion of pigment granules, while during aggregation, myosin V is switched off, presumably by phosphorylation-dependent release from the granules⁶⁰. In vertebrate melanophores⁶¹ or neurons⁶², the two classes of motors may act sequentially. Fast, long-range microtubule-dependent transport in the cell body is supplanted by short-range actin-dependent transport in the cell periphery. Here myosin V, with its long neck and large stride, may safely haul its cargo through the tangle of cortical actin filaments, not unlike an ape swaying from branch to branch in a treetop. These examples suggest the intriguing possibility that the deployment of many cell organelles depends on the concerted action of multiple motors. Myosin V and conventional kinesin have been shown to interact directly in their tail domains⁶³, but it remains to be seen whether physical interaction of motors is the key to their coordination.

Motors in novel contexts

Organelle transport and ciliary movement or contraction are paradigmatic tasks of cytoskeletal motors, but there is more to motors

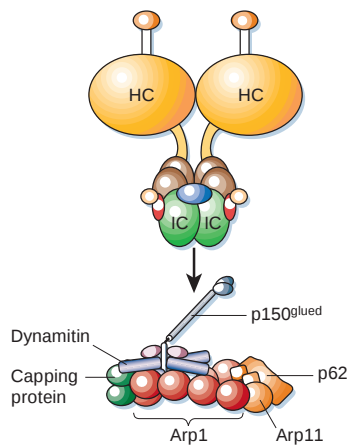


Figure 4 Schematic overview of the dynein–dynactin complex. The dynein molecule, itself a complex of heavy (HC), intermediate (IC) and light chains, interacts with the p150^{glued} subunit of the dynactin complex through its intermediate chains (arrow), although the precise mode of interaction is not known. The most prominent component of the dynactin complex is a short filament of the actin-related protein Arp1.

than meets the microscopist's eye. Some motors are implicated in the transport of messenger RNA or macromolecular complexes. Others are unable to move and yet are indispensable for certain cellular activities (see review in this issue by Howard and Hyman, page 753). Some deletions or mutations of motors can be lethal for multicellular organisms, indicating that these motors are essential for crucial steps in development. In other circumstances, the loss of certain motors leads to debilitating diseases. Finally, motors may participate in cellular homeostasis and cell architecture in ways that extend beyond functions in transport. These are exciting research fields unforeseen only a few years ago.

New on the agenda of motor functions is an involvement in mRNA transport. Restricting mRNA translation restricts the subcellular distribution of the protein product but requires the transport of mRNA to its destination. Depending on the system studied, mRNA transport is accomplished by myosin, kinesin or dynein motors. In yeast, for example, certain mRNAs are transported in a complex with myosin V (ref. 64), whereas in neurons or insect oocytes, microtubule motors are required^{65,66}. In all cases, RNA-binding and adaptor proteins integrate the RNA into a ribonucleoprotein transport package, although the precise molecular interactions within this complex have yet to be determined. A paradigm for the extraordinary importance of motor-dependent mRNA localization is the *Drosophila* oocyte where the convergence of *oskar* mRNA and associated proteins at the posterior pole is supported by conventional kinesin^{67,68}, whereas *bicoid* mRNA-containing complexes are moved by dynein to the anterior pole⁶⁹. Their precise deployment establishes the anterior–posterior axis. There are hints that the establishment of dorsoventrality requires motors as well.

The determination of the left–right axis in mammals also depends on the activity of molecular motors, although the nature of the implication is different. Left–right patterning was suggested to require the transport of a 'morphogen' to the left side of the vertebrate gastrula by cilia of the embryonic node, an organizing structure in the developing embryo. In support of this notion, mutations in a gene encoding a dynein isoform known as left–right dynein⁷⁰ and in a member of the Krp85/95 kinesin family⁷¹, both of which are required for ciliary development, inactivate nodal cilia and lead to random positioning of internal organs. The conclusions from these studies met with scepticism, because the mechanism was believed to be too

unspecific for such a crucial step in development. However, an artificial flow around the nodal cilia, generated with the use of an ingenious micromechanical device, was shown to influence the positioning of internal organs⁷².

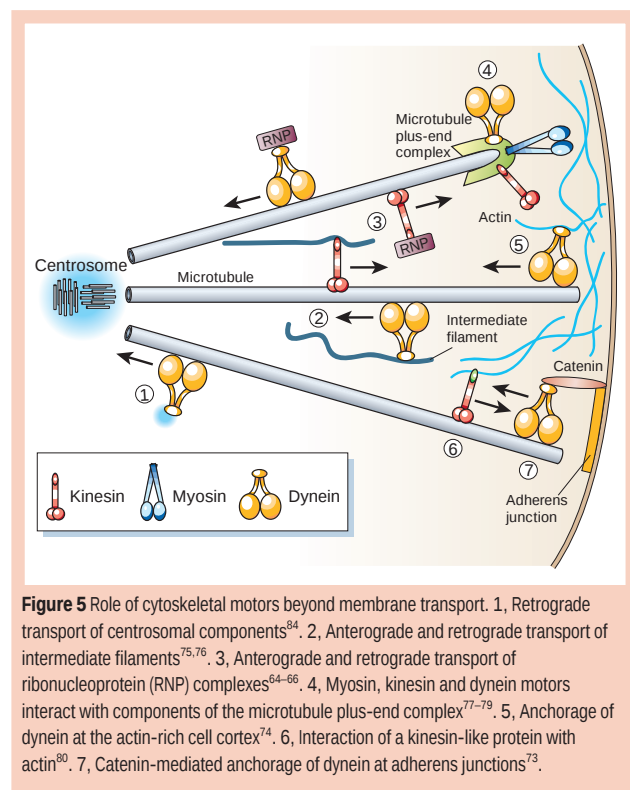
This is just one striking example that a motor defect can lead to a pathological condition, situs inversus. Work over the past few years has implicated motors in a growing number of human diseases (Table 1). So far, these can be grouped into five categories: defects associated with contraction that can be traced to myosin II; sensory defects associated with several unconventional myosins; disorders associated with defects in ciliary biogenesis and function that are linked to axonemal dynein and the kinesin-like protein Krp85/95, as described above; intracellular transport deficiencies attributable to defects in cytoplasmic dynein and kinesin organelle transporters; and transport of pathogens. These implications are firmly established in some cases (for example, myopathies or hearing loss) and more tenuous in others (for example, neurodegenerative diseases), but they are tantalizing enough to spur further efforts aimed at the discovery of hidden links between motors and disease.

A final issue concerns an involvement of motors in cell architecture and cytoskeletal remodelling (Fig. 5). We have seen that some motors can be part of large macromolecular complexes and, through their associated proteins, can interact with a wide spectrum of cytoplasmic constituents. A paradigm is the dynein/dynactin machinery, which has been shown to be important not only in organelle transport, but also in cytoskeletal architecture. Dynein associates with adherens junctions of epithelial cells through an interaction with β -catenin and a novel protein, PLAC-24, that binds the dynein intermediate chain⁷³. This protein complex may help to tether microtubule ends at sites of cell–cell contact. Additional interactions of cortical cytoplasmic dynein with microtubule plus ends affect spindle orientation, nuclear movement, centrosome positioning and cell polarity⁷⁴. Thus, cortical dynein can profoundly influence the spatial organization of the entire microtubule apparatus, which in turn provides a framework for the organization of cellular membrane systems. In addition, dynein as well as conventional kinesin are required for the assembly and dynamics of the vimentin intermediate filament system⁷⁵ and neurofilament transport⁷⁶, supporting the long-standing notion of a close spatial relationship between these two cytoskeletal systems. Both motors, or their interacting proteins, can also be part of the microtubule plus-end complex, a large assemblage of proteins associated with growing microtubule ends^{77,78}.

Motors may link the microtubule and actin systems as well. For example, a class VI myosin interacts with a microtubule plus-end-binding protein⁷⁹; CHO1, a kinesin of the MKLP1 subfamily, possesses an extra domain that interacts with actin filaments⁸⁰; actin reorganization requires a ras-related GTPase that interacts specifically with KIF9 kinesin⁸¹; a dynein light chain of relative molecular mass 8,000 (DLC8) is also a component of myosin V⁸²; and a plant kinesin, KCBP, possesses a myosin tail homology domain, a widespread subdomain of several myosins⁸³. Sporadic as they may seem, these findings hint at a system of functional interactions between cytoskeletal systems mediated by molecular motors. Coupled with the observation that microtubule motors help construct large-scale assemblies such as centrosomes⁸⁴ or microtubule asters⁸⁵, and that myosin tunes viscoelasticity without disrupting filament networks⁸⁶, motors are emerging as dynamic modulators of cell architecture.

Outlook

Extrapolating into the future is always challenging and often wrong. Using current work as a guide, four main areas of future research on molecular motors can be identified. First, even though we seem to have a general idea of motor chemomechanics, important details still need to be worked out. Atomic resolution structures will be the guide. In combination with single-molecule techniques of improved spatiotemporal resolution and sensitivity and the rational design of



motor mutants, common principles of motor physiology will emerge. Second, many motors are known only by sequence, particularly in plants, so this is a fertile playground for the cell biological hunter-gatherer. Functional characterization will help answer questions of motor targeting and motor regulation: how does a motor find its cargo, what directs it to the correct target site, and how is its activity regulated in the process? Only partial answers are available at present. Third, the implication of motors in disease and developmental defects will attract increasing attention. The questions, and the answers they demand, will undoubtedly be complex, as motor defects will frequently be just one of many factors that contribute to the manifestation of a disease. Fourth, motors are believed to hold promise for use in nanobiotechnological devices, although marketable applications have yet to be achieved.

“Our progress is narrow; it takes a vast world unchallenged and for granted,” writes J. Robert Oppenheimer⁸⁷. “This is why we will have to accept the fact that no one of us really will ever know very much. This is why we shall have to find comfort in the fact that, taken together, we know more and more.” The field of molecular motors is no exception. □

doi:10.1038/nature01601

- Schliwa, M. (ed.) *Molecular Motors* (VCH-Wiley, Weinheim, 2003).
- Vale, R. D. Switches, latches, and amplifiers: common themes of G proteins and molecular motors. *J. Cell Biol.* **135**, 291–302 (1996).
- Kosztin, I., Bruinsma, R., O’Lague, P. & Schulten, K. Mechanical force generation by G proteins. *Proc. Natl Acad. Sci. USA* **99**, 3575–3580 (2002).
- Yun, M., Zhang, X., Park, C. G., Park, H. W. & Endow, S. A. A structural pathway for activation of the kinesin motor ATPase. *EMBO J.* **20**, 2611–2618 (2000).
- Murphy, C. T., Rock, R. S. & Spudich, J. A. A myosin II mutation uncouples ATPase activity from motility and shortens step size. *Nature Cell Biol.* **3**, 311–315 (2002).
- Geeves, M. A. & Holmes, K. C. Structural mechanism of muscle contraction. *Annu. Rev. Biochem.* **68**, 687–728 (1999).
- Houdusse, A., Szent-Gyorgyi, A. G. & Cohen, C. Three conformational states of scallop myosin S1. *Proc. Natl Acad. Sci. USA* **21**, 11238–11243 (2000).
- Spudich, J. A. The myosin swinging cross-bridge model. *Nature Rev. Mol. Cell Biol.* **2**, 387–392 (2001).
- Ruff, C., Furch, M., Brenner, B., Manstein, D. J. & Meyhofer, E. Single-molecule tracking of myosins with genetically engineered amplifier domains. *Nature Struct. Biol.* **8**, 226–229 (2001).
- Schott, D. H., Collins, R. N. & Bretscher, A. Secretory vesicle transport velocity in living cells depends on the myosin-V lever arm length. *J. Cell Biol.* **156**, 35–39 (2002).

- Rice, S. et al. A structural change in the kinesin motor protein that drives motility. *Nature* **402**, 778–784 (1999).
- Kikkawa, M., Okada, Y. & Hirokawa, N. 15 Å resolution model of the monomeric kinesin motor, KIF1A. *Cell* **100**, 241–252 (2000).
- Vale, R. D. & Milligan, R. A. The way things move: looking under the hood of molecular motor proteins. *Science* **288**, 88–95 (2000).
- Koonce, M. P. & Tikhonenko, I. Functional elements within the dynein microtubule-binding domain. *Mol. Biol. Cell* **11**, 523–529 (2000).
- Burgess, S. A., Walker, M. L., Sakakibara, H., Knight, P. J. & Oiwa, K. Dynein structure and power stroke. *Nature* **421**, 715–718 (2003).
- Hackney, D. D. Highly processive microtubule-stimulated ATP hydrolysis by dimeric kinesin head domains. *Nature* **377**, 448–450 (1995).
- Hua, W., Chung, J. & Gelles, J. Distinguishing inchworm and hand-over-hand processive kinesin movement by neck rotation measurements. *Science* **295**, 844–848 (2002).
- Kozlowski, F. et al. The crystal structure of dimeric kinesin and implications for microtubule-dependent motility. *Cell* **91**, 985–994 (1997).
- Kallipoliti, A. et al. Unusual properties of the fungal conventional kinesin neck domain from *Neurospora crassa*. *EMBO J.* **20**, 6226–6235 (2001).
- Tomishige, M. & Vale, R. D. Controlling kinesin by reversible disulfide cross-linking. Identifying the motility-producing conformational change. *J. Cell Biol.* **151**, 1081–1092 (2000).
- Rief, M. et al. Myosin-V stepping kinetics: a molecular model for processivity. *Proc. Natl Acad. Sci. USA* **97**, 9482–9486 (2000).
- De La Cruz, E. M., Wells, A. L., Rosenfeld, S. S., Ostap, E. M. & Sweeney, H. L. The kinetic mechanism of myosin V. *Proc. Natl Acad. Sci. USA* **96**, 13726–13731 (1999).
- Veigel, C., Wang, F., Bartoo, M. L., Sellers, J. R. & Molloy, J. E. The gated gait of the processive molecular motor, myosin V. *Nature Cell Biol.* **4**, 59–65 (2002).
- Rock, R. S. et al. Myosin VI is a processive motor with a large step size. *Proc. Natl Acad. Sci. USA* **98**, 13655–13659 (2001).
- Walker, M. L. et al. Two-headed binding of a processive myosin to F-actin. *Nature* **405**, 804–807 (2000).
- Okada, Y. & Hirokawa, N. A processive single-headed motor: kinesin superfamily protein KIF1A. *Science* **283**, 1152–1157 (1999).
- Inoue, A., Saito, J., Ikebe, R. & Ikebe, M. Myosin IXb is a single-headed minus-end-directed processive motor. *Nature Cell Biol.* **4**, 302–306 (2002).
- Sakakibara, H., Kojima, H., Saka, Y., Katayama, E. & Oiwa, K. Inner-arm dynein c of *Chlamydomonas* flagella is a single-headed processive motor. *Nature* **400**, 586–590 (1999).
- Okada, Y. & Hirokawa, N. Mechanism of the single-headed processivity: diffusional anchoring between the K-loop of kinesin and the C terminus of tubulin. *Proc. Natl Acad. Sci. USA* **97**, 640–645 (2000).
- Thorn, K. S., Ubersax, J. A. & Vale, R. D. Engineering the processive run length of the kinesin motor. *J. Cell Biol.* **151**, 1093–1100 (2000).
- Rogers, K. R. et al. KIF1D is a fast non-processive kinesin that demonstrates novel K-loop-dependent mechanochemistry. *EMBO J.* **20**, 5101–5113 (2001).
- Tomishige, M., Klopfenstein, D. R. & Vale, R. D. Conversion of Unc104/KIF1A kinesin into a processive motor after dimerization. *Science* **297**, 2263–2267 (2002).
- Sablin, E. P. et al. Direction determination in the minus-end-directed kinesin motor ncd. *Nature* **395**, 813–816 (1998).
- Endow, S. A. Determinants of molecular motor directionality. *Nature Cell Biol.* **1**, E163–E167 (1999).
- Endow, S. A. & Higuchi, H. A mutant of the motor protein kinesin that moves in both directions on microtubules. *Nature* **406**, 913–916 (2000).
- Wells, A. L. et al. Myosin VI is an actin-based motor that moves backwards. *Nature* **401**, 505–508 (1999).
- Homma, K., Yoshimura, M., Saito, J., Ikebe, R. & Ikebe, M. The core of the motor domain determines the direction of myosin movement. *Nature* **412**, 831–834 (2001).
- Howard, J. *Mechanics of Motor Proteins and the Cytoskeleton* (Sinauer, Sunderland, MA, 2001).
- Schnitzer, M. J., Visscher, K. & Block, S. M. Force production by single kinesin motors. *Nature Cell Biol.* **2**, 718–723 (2000).
- Lang, M. J., Asbury, C. L., Shaevitz, J. W. & Block, S. M. An automated two-dimensional optical force clamp for single molecule studies. *Biophys. J.* **83**, 491–501 (2002).
- Tang, N., Lin, T. & Ostap, E. M. Dynamics of myo1c (myosin-1β) lipid binding and dissociation. *J. Biol. Chem.* (2002).
- Klopfenstein, D. R., Tomishige, M., Stuurman, N. & Vale, R. D. Role of phosphatidylinositol(4,5)bisphosphate organization in membrane transport by the Unc104 kinesin motor. *Cell* **109**, 347–358 (2002).
- Kamal, A. et al. Kinesin-mediated axonal transport of a membrane compartment containing β-secretase and presenilin-1 requires APP. *Nature* **414**, 643–648 (2001).
- Tai, A. W., Chuang, J. Z., Bode, C., Wolfgram, U. & Sung, C. H. Rhodopsin’s carboxy-terminal cytoplasmic tail acts as a membrane receptor for cytoplasmic dynein by binding to the dynein light chain Tctex-1. *Cell* **97**, 877–887 (1999).
- Bowman, A. B. et al. Kinesin-dependent axonal transport is mediated by the sunday driver (SYD) protein. *Cell* **10**, 583–594 (2000).
- Verhey, K. J. et al. Cargo of kinesin identified as JIP scaffolding proteins and associated signaling molecules. *J. Cell Biol.* **152**, 959–970 (2001).
- Terada, S. & Hirokawa, N. Moving on to the cargo problem of microtubule-dependent motors in neurons. *Curr. Opin. Neurobiol.* **10**, 566–573 (2000).
- Wu, X. S. et al. Identification of an organelle receptor for myosin-Va. *Nature Cell Biol.* **4**, 271–278 (2002).
- Hammer, J. A. III & Wu, X. S. Rabs grab motors: defining the connections between Rab GTPases and motor proteins. *Curr. Opin. Cell Biol.* **14**, 69–75 (2002).
- Hollerer, E. A., Karki, S. & Holzbaur, E. L. The role of the dynactin complex in intracellular motility. *Int. Rev. Cytol.* **18**, 69–109 (1998).
- Muresan, V. et al. Dynactin-dependent, dynein-driven vesicle transport in the absence of membrane proteins: a role for spectrin and acidic phospholipids. *Mol. Cell* **7**, 173–183 (2001).
- Karcher, R. L. et al. Cell cycle regulation of myosin-V by calcium/calmodulin-dependent protein kinase II. *Science* **293**, 1317–1320 (2001).

53. Addinall, S. G. *et al.* Phosphorylation by cdc2-cyclinB1 kinase releases cytoplasmic dynein from membranes. *J. Biol. Chem.* **276**, 15939–15944 (2001).
54. Morfini, G., Szebenyi, G., Elluru, R., Ratner, N. & Brady, S. T. Glycogen synthase kinase 3 phosphorylates kinesin light chains and negatively regulates kinesin-based motility. *EMBO J.* **21**, 281–293 (2002).
55. Coy, D. L., Hancock, W. O., Wagenbach, M. & Howard, J. Kinesin's tail domain is an inhibitory regulator of the motor domain. *Nature Cell Biol.* **1**, 288–292 (1999).
56. Friedman, D. S. & Vale, R. D. Single-molecule analysis of kinesin motility reveals regulation by the cargo-binding tail domain. *Nature Cell Biol.* **1**, 293–297 (1999).
57. Seiler, S. *et al.* Cargo binding and regulatory sites in the tail of fungal conventional kinesin. *Nature Cell Biol.* **2**, 333–338 (2000).
58. Hackney, D. D. & Stock, M. F. Kinesin's IAK tail domain inhibits initial microtubule-stimulated ADP release. *Nature Cell Biol.* **2**, 257–260 (2000).
59. Brown, S. S. Cooperation between microtubule- and actin-based motor proteins. *Annu. Rev. Cell Dev. Biol.* **15**, 63–80 (1999).
60. Gross, S. P. *et al.* Interactions and regulation of molecular motors in *Xenopus* melanophores. *J. Cell Biol.* **156**, 855–865 (2002).
61. Wu, X. S. *et al.* Identification of an organelle receptor for myosin-Va. *Nature Cell Biol.* **4**, 271–278 (2002).
62. Bridgman, P. C. Myosin Va movements in normal and dilute-lethal axons provide support for a dual filament motor complex. *J. Cell Biol.* **146**, 1045–1060 (1999).
63. Huang, J. D. *et al.* Direct interaction of microtubule- and actin-based transport motors. *Nature* **397**, 267–270 (1999).
64. Jansen, R. P. mRNA localization: message on the move. *Nature Rev. Mol. Cell Biol.* **2**, 247–256 (2001).
65. Kiebler, M. A. & DesGroseillers, L. Molecular insights into mRNA transport and local translation in the mammalian nervous system. *Neuron* **25**, 19–28 (2000).
66. Stebbings, H. Cytoskeleton-dependent transport and localization of mRNA. *Int. Rev. Cytol.* **211**, 1–31 (2001).
67. Brendza, R. P., Serbus, L. R., Duffy, J. B. & Saxton, W. M. A function for kinesin I in the posterior transport of *oskar* mRNA and Staufen protein. *Science* **289**, 2120–2122 (2000).
68. Cha, B. J., Koppetsch, B. S. & Theurkauf, W. E. In vivo analysis of *Drosophila bicoid* mRNA localization reveals a novel microtubule-dependent axis specification pathway. *Cell* **106**, 35–46 (2001).
69. Schnorrer, F., Bohmann, K. & Nüsslein-Volhard, C. The molecular motor dynein is involved in targeting swallow and bicoid RNA to the anterior pole of *Drosophila* oocytes. *Nature Cell Biol.* **2**, 185–190 (2000).
70. Supp, D. M., Witte, D. P., Potter, S. S. & Brueckner, M. Mutation of an axonemal dynein affects left–right asymmetry in *inversus viscerum* mice. *Nature* **389**, 963–966 (1997).
71. Nonaka, S. *et al.* Randomization of left–right asymmetry due to loss of nodal cilia generating leftward flow of extraembryonic fluid in mice lacking KIF3B motor protein. *Cell* **95**, 829–837 (1998).
72. Nonaka, S., Shiratori, H., Saijoh, Y. & Hamada, H. Determination of left–right patterning of the mouse embryo by artificial nodal flow. *Nature* **418**, 96–99 (2002).
73. Karki, S., Ligon, L. A., DeSantis, J., Tokito, M. & Holzbaur, E. L. PLAC-24 is a cytoplasmic dynein-binding protein that is recruited to sites of cell–cell contact. *Mol. Biol. Cell* **13**, 1722–1734 (2002).
74. Dujardin, D. L. & Vallee, R. B. Dynein at the cortex. *Curr. Opin. Cell Biol.* **14**, 44–49 (2002).
75. Helfand, B. T., Mikami, A., Vallee, R. B. & Goldman, R. D. A requirement for cytoplasmic dynein and dynactin in intermediate filament network assembly and organization. *J. Cell Biol.* **157**, 795–806 (2002).
76. Shea, T. B. & Flanagan, L. A. Kinesin, dynein and neurofilament transport. *Trends Neurosci.* **24**, 644–648 (2001).
77. Jimbo, T. *et al.* Identification of a link between the tumour suppressor APC and the kinesin superfamily. *Nature Cell Biol.* **4**, 323–327 (2002).
78. Askham, J. M., Vaughan, K. T., Goodson, H. V. & Morrison, E. E. Evidence that an interaction between EB1 and p150^{Glued} is required for the formation and maintenance of a radial microtubule array anchored at the centrosome. *Mol. Biol. Cell* **13**, 3627–3645 (2002).
79. Lantz, V. A. & Miller, K. G. A class VI unconventional myosin is associated with a homologue of a microtubule-binding protein, cytoplasmic linker protein-170, in neurons and at the posterior pole of *Drosophila* embryos. *J. Cell Biol.* **140**, 897–910 (1999).
80. Kuriyama, R., Gustus, C., Terada, Y., Uetake, Y. & Matuliene, J. CHO1, a mammalian kinesin-like protein, interacts with F-actin and is involved in the terminal phase of cytokinesis. *J. Cell Biol.* **156**, 783–790 (2002).
81. Piddini, E., Schmid, J. A., de Martin, R. & Dotti, C. G. The Ras-like GTPase Gem is involved in cell shape remodelling and interacts with the novel kinesin-like protein KIF9. *EMBO J.* **20**, 4076–4087 (2001).
82. Espindola, F. S. *et al.* The light chain composition of chicken brain myosin-Va: calmodulin, myosin-II essential light chains, and 8-kDa dynein light chain/PIN. *Cell Motil. Cytoskel.* **47**, 269–281 (2000).
83. Day, I. S., Miller, C., Golovkin, M. & Reddy, A. S. Interaction of a kinesin-like calmodulin-binding protein with a protein kinase. *J. Biol. Chem.* **275**, 13737–13745 (2000).
84. Zimmerman, W. & Dosey, S. J. Construction of centrosomes and spindle poles by molecular motor-driven assembly of protein particles. *Traffic* **1**, 927–934 (2000).
85. Surrey, T., Nedelec, F., Leibler, S. & Karsenti, E. Physical properties determining self-organization of motors and microtubules. *Science* **292**, 1167–1171 (2001).
86. Humphrey, D., Duggan, C., Saha, D., Smith, D. & Käs, J. Active fluidization of polymer networks through molecular motors. *Nature* **416**, 413–416 (2002).
87. Oppenheimer, J. R. *Science and the Common Understanding* (Simon & Schuster, 1953).
88. Seidman, J. G. & Seidman, C. The genetic basis for cardiomyopathy: from mutation identification to mechanistic paradigms. *Cell* **104**, 557–567 (2001).
89. Westbroek, W., Lambert, J. & Naeyaert, J. M. The *dilute* locus and Griscelli syndrome: gateways towards a better understanding of melanosome transport. *Pigment Cell Res.* **14**, 320–327 (2001).
90. Avraham, K. B. The genetics of deafness: a model for genomic and biological complexity. *Ernst Schering Res. Found. Workshop* **36**, 271–297 (2002).
91. Marszalek, J. R. *et al.* Genetic evidence for selective transport of opsin and arrestin by kinesin-II in mammalian photoreceptors. *Cell* **102**, 175–187 (2000).
92. Olbrich, H. *et al.* Mutations in *DNAH5* cause primary ciliary dyskinesia and randomization of left–right asymmetry. *Nature Genet.* **30**, 143–144 (2002).
93. Marszalek, J. R., Ruiz-Lozano, P., Roberts, E., Chien, K. R. & Goldstein, L. S. Situs inversus and embryonic ciliary morphogenesis defects in mouse mutants lacking the KIF3A subunit of kinesin-II. *Proc. Natl Acad. Sci. USA* **96**, 5043–5048 (1999).
94. Qin, H., Rosenbaum, J. L. & Barr, M. M. An autosomal recessive polycystic kidney disease gene homolog is involved in intraflagellar transport in *C. elegans* ciliated sensory neurons. *Curr. Biol.* **11**, 457–461 (2001).
95. Vallee, R. B., Tai, C. & Faulkner, N. E. LIS1: cellular function of a disease-causing gene. *Trends Cell Biol.* **11**, 155–160 (2001).
96. Zhao, C. *et al.* Charcot-Marie-Tooth disease type 2A caused by mutation in a microtubule motor KIF1B. *Cell* **105**, 587–597 (2001).
97. Rietdorf, J. *et al.* Kinesin-dependent movement on microtubules precedes actin-based motility of vaccinia virus. *Nature Cell Biol.* **3**, 992–1000 (2001).
98. Dohner, K. *et al.* Function of dynein and dynactin in herpes simplex virus capsid transport. *Mol. Biol. Cell* **13**, 2795–2809 (2002).
99. Watters, J. W., Dewar, K., Lehoczy, J., Boyartchuk, V. & Dietrich, W. F. Kif1C, a kinesin-like motor protein, mediates mouse macrophage resistance to anthrax lethal factor. *Curr. Biol.* **11**, 1503–1511 (2001).
100. Gunawardena, S. & Goldstein, L. S. Disruption of axonal transport and neuronal viability by amyloid precursor protein mutations in *Drosophila*. *Neuron* **32**, 389–401 (2001).
101. LaMonte, B. H. *et al.* Disruption of dynein/dynactin inhibits axonal transport in motor neurons causing late-onset progressive degeneration. *Neuron* **34**, 715–727 (2002).

Acknowledgements Space limitations imposed a strict limit on the number of references. We apologize to all authors who made significant contributions to the field but whose work is not cited. Our work is supported by the Deutsche Forschungsgemeinschaft and the Fonds der Chemischen Industrie.

Adaptation of core mechanisms to generate cell polarity

W. James Nelson

Department of Molecular and Cellular Physiology, Beckman Center for Molecular and Genetic Medicine, Stanford University School of Medicine, Stanford, California 94305-5435, USA (e-mail: wjnelson@stanford.edu)

Cell polarity is defined as asymmetry in cell shape, protein distributions and cell functions. It is characteristic of single-cell organisms, including yeast and bacteria, and cells in tissues of multi-cell organisms such as epithelia in worms, flies and mammals. This diversity raises several questions: do different cell types use different mechanisms to generate polarity, how is polarity signalled, how do cells react to that signal, and how is structural polarity translated into specialized functions? Analysis of evolutionarily diverse cell types reveals that cell-surface landmarks adapt core pathways for cytoskeleton assembly and protein transport to generate cell polarity.

Most cells are polarized, including simple single-cell organisms (for example, budding yeast *Saccharomyces cerevisiae* and fission yeast *Schizosaccharomyces pombe*) and cells in multicellular invertebrates (the nematode *Caenorhabditis elegans* or fruitfly *Drosophila*) and vertebrates (mammals); even bacteria are polarized (Fig. 1). There is also an extraordinary diversity in the shapes of polarized cell. Consider the extremely attenuated shape of a mammalian neuron, which can be many metres long, the short rectangular shapes (5–30 μm long) of the single-cell *S. pombe* and cells in epithelial tissues in multicellular organisms, and the asymmetric shape of *S. cerevisiae* or *Drosophila* neuroblasts preparing to divide (Fig. 1). Different cell shapes are not just a quirk of nature, but are coupled to specialized cell functions, for example, to communicate between tissues over long distances (neurons), to provide barriers that regulate ionic homeostasis between different biological compartments (epithelia), and to disperse cellular components to daughter cells upon cell division.

At first glance, this diversity of cell shapes and functions suggests that each cell type must have evolved completely different ways to generate cell polarity that distinguishes, for example, a budding yeast from a multi-cell epithelium. But although the final structure is quite different, the basic toolbox of core mechanisms used to organize the cytoskeleton and deliver membrane proteins is common to all eukaryotic cell types. This review focuses on how these mechanisms are used to generate polarity in evolutionarily diverse cell types with different shapes and functions.

Generating polarity in single cells for mitotic division

S. cerevisiae and *S. pombe* are simple, well-understood examples of how cells generate polarity and couple it to a specific function; simple because they are single-cell organisms that generate cell polarity in order to divide, and well understood because they are genetically tractable. Genetic and cell biological studies point to the importance of localized assembly of actin patches and regulatory proteins to mark one site on the mother cell for membrane growth, and the attachment of actin cables and microtubules to those sites for vesicle delivery and spindle orientation, respectively.

During vegetative growth, *S. cerevisiae* adopt genotype-dependent axial (haploid α or α cells) or bipolar (diploid α/α cells) growth patterns to produce a daughter cell bud¹ (Fig. 2a). Two classes of genes prescribe 'landmarks' for axial (*BUD3*, *BUD4* and *BUD10/Axl2*) and bipolar (*BUD8* and *BUD9*) budding of the daughter cell^{2,3}, and a third class is required for both budding patterns (*BUD1/RSR1*, *BUD2* and *BUD5*)^{2,4}. The landmark is retained on the plasma membrane close to (axial or bipolar pattern) or opposite (bipolar pattern) the site (bud scar) of the preceding cytokinesis (Fig. 2a). Membrane growth is anisotropic, occurring only in the new bud (Fig. 2b).

Members of the Ras superfamily of small GTPases (*Bud1p/Rsr1p*, *Cdc42p*, and *Rho1p/2p*, *Rho3p* and *Rho4p*) coordinate the selection of the membrane growth site (the bud), and orientation of the cytoskeleton and targeting of vesicles to that site^{1,5} (Fig. 2b, c). *Cdc42p* is crucial in these events — in the absence *Cdc42p* the actin cytoskeleton is disorganized, and large, unbudded cells are formed as a result of isotropic growth of the mother cell⁶. A critical question is how is *Cdc42p* activity restricted to one site on the plasma membrane? One possibility is that *Cdc42p* is restricted to the landmark through the sequential recruitment of guanine-exchange factors (GEFs), which control GTPase activity locally⁷. *Bud5p* and *Bud2p*, the GEF and GTPase-activating protein for *Bud1p/Rsr1p*, respectively, co-localize with landmark proteins, and disruption of landmark proteins results in mislocalization of *Bud5p*^{8–10}. Recruitment of *Bud5p* to the landmark would locally activate the uniformly distributed *Bud1p/Rsr1p*. *In vitro* studies¹¹ show that the GTP-bound form of *Bud1p/Rsr1p* binds *Cdc42p*, the GEF for *Cdc42p*, which in the cell could restrict *Cdc42p* activity to the landmark. But cells lacking the *Bud1p/Rsr1p* machinery or any of the landmark genes form daughter cell buds, albeit randomly on the mother cell, in a process that requires activated *Cdc42p*^{2–4}. Polarization of *Cdc42p* and actin polymerization could be controlled through a positive feedback loop initiated by a stochastic increase in *Cdc42p* on the membrane¹², and landmarks normally adapt this feedback loop for bud formation at a prescribed, rather than random, site on the membrane.

Activation of *Cdc42p* at a site on the membrane unleashes global changes in cytoskeleton organization within the bud, involving assembly of a cap of actin filaments, and the

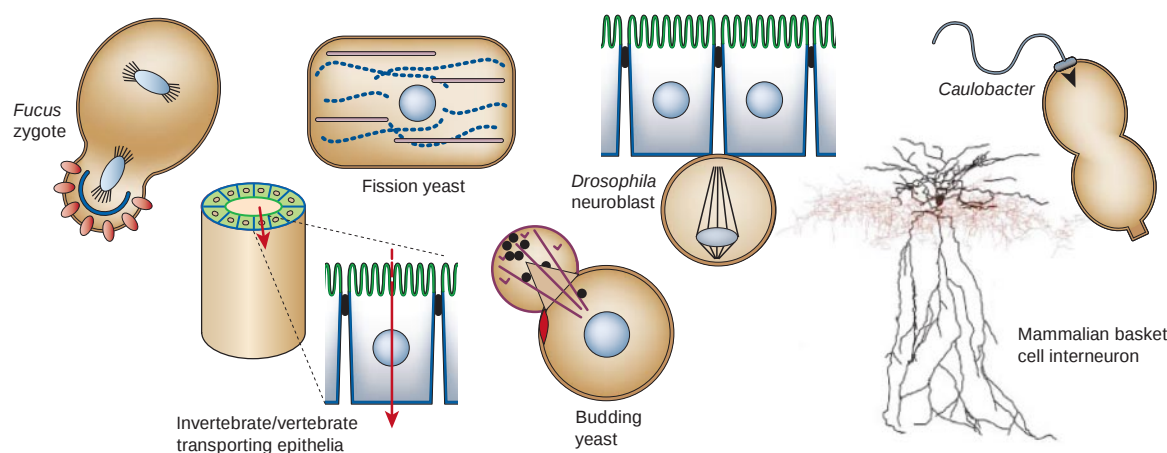


Figure 1 Diversity of shapes of polarized cells (not to scale). *Fucus* zygote exposed to a light gradient showing polarized distributions of ion channels/dihydropyridine receptors (red circles) and F-actin (blue line) in the rhizoid cell (bottom) compared to the thallus cell (top). Fission yeast (*Schizosaccharomyces pombe*) showing polarized distributions of actin (purple) and microtubules (blue dotted line) in the long axis of the cell, and the nucleus positioned in the centre of the cell. *Drosophila* neuroblast delaminated from the ventral neuroectoderm with an asymmetric plane of division that will yield a large 'apical' neuroblast stem cell and a small 'basal' ganglion mother cell. *Caulobacter crescentus* predivisional cell showing polarized distributions of the flagellum (swarmer cell, top) and stalk (stalked cell, bottom). Invertebrate/vertebrate transporting epithelium showing organization of polarized epithelial cells (apical membrane, green; basolateral membrane, blue) in a tube that separates two biological compartments and regulates vectorial transport of ions/solutes (red arrow) between those compartments. Budding yeast (*Saccharomyces cerevisiae*) forming a daughter cell 'bud' from the mother cell next to the previous site of cytokinesis (bud scar, red disc), and orienting actin cables (purple) for transport of vesicles (black circles) from the mother to daughter cell. Mammalian basket cell interneuron showing the distribution of the soma/dendrite (black) and axon (red; image courtesy of D. Madison, Stanford University School of Medicine).

organization of actin cables that extend into the mother cell^{1,13} (Fig. 2b, c). Ste20p and Cla4p are downstream effectors of Cdc42p that are members of the p21-activated kinase (PAK) family¹⁴. The type-I myosins Myo3p and Myo5p¹⁵ are PAK substrates, which together with homologues of Wiskott–Aldrich syndrome protein or WASP (Bee1p/Las17p) and WASP-interacting protein or WIP (Vrp1p) interact with the Arp2/3 complex to assemble and organize an actin cap at the bud site^{16–18} (for details of Arp2/3 function, see review in this issue by Gruenheid and Finlay, page 775).

A core scaffolding complex of Spa2p, Pea2p and Sph1p (the polarisome complex¹⁹) is assembled at the bud. The formin homologues Bni1p and Bnr1p bind this complex and, in turn, interact with proteins that assemble actin cables between the bud and mother cell^{20–22}: profilin (Pfy1p)²³, which promotes GDP to GTP exchange in monomeric actin and hence provides a local pool of GTP-actin for addition to the barbed end of growing actin filaments; Bud6p/Aip3p, an actin filament-binding protein²⁴; and activated GTPases including Cdc42p, Rho1p/2p, Rho3p and Rho4p^{20,25}. Loss of Bni1p and Bnr1p activity correlates with loss of actin cables throughout the bud and mother, and re-activation of Bni1p and Bnr1p results in re-growth of actin cables from the bud into the mother cell^{21,22}. *In vitro*, Bni1p can also nucleate actin polymerization and bind the barbed (growing) end of actin filaments^{26,27}. Taken together, Bni1p and Bnr1p and associated proteins seem to nucleate and anchor actin cables at the bud site (Fig. 2c).

Orientation of actin cables from the bud into the mother cell also drives microtubule capture in the bud and spindle orientation in the mother–bud axis for cytokinesis²⁸ (Fig. 2b). Astral microtubules contain at their distal, plus ends a protein complex of Kar9p and Bim1p that binds to the type-V myosin Myo2p, which, in turn, binds to actin cables and moves the microtubule-attached complex into the bud²⁹.

Actin cables between the bud and mother cell provide tracks for exocytic vesicle transport from the mother cell Golgi complex into the bud^{1,13} (Fig. 2b, c). This is critical for polarized membrane growth as SNAREs (soluble N-ethylmaleimide-sensitive factor attachment protein receptors) located on the target membrane (t-SNAREs; Sec9p, Sso1p/2p) are not restricted to the bud, but are localized over

the surface of the mother cell and bud³⁰. Vesicle delivery along actin cables requires Myo2p^{31,32}. Vesicle docking/fusion complex at the bud is further specified by a large complex termed the exocyst, which is localized initially to the bud tip³³; Sec4p, Rho1p/2p, Rho3p, Rho4p and Cdc42p regulate these late stages in vesicle delivery, and localization and function of the exocyst in the bud^{5,34}. Disruption of genes encoding the exocyst complex results in accumulation of transport vesicles in the bud, indicating that the exocyst complex may tether arriving vesicles or activate the SNARE complex in preparation for SNARE-dependent vesicle fusion with the plasma membrane³³. The yeast homologues of *Drosophila* lethal giant larva (a protein regulating epithelial polarity, see below), *sro7/sro77*, bind the t-SNARE Sec9p, and *sro7/sro77* mutants accumulate post-Golgi transport vesicles in the cytoplasm³⁵. Thus, both t-SNAREs and the exocyst are required for vesicle fusion at the bud, and localized vesicle docking/fusion in the bud is specified by vesicle delivery along actin cables to the exocyst.

Although these overlapping regulatory mechanisms involving different GTPases may seem overly complex, and in some cases redundant, they coordinate a series of processes that link, in time and space, the initial selection of the bud site to localized assembly and orientation of the actin and microtubule cytoskeletons in the bud–mother axis, and localized docking/fusion of vesicles at that site of membrane growth (Fig. 2c).

Compared to the highly polarized bud growth of *S. cerevisiae*, the rod-shaped fission yeast *S. pombe* divides in the middle by septation and, therefore, might be expected to grow in a more isotropic manner during the cell cycle. However, growth of fission yeast is also spatially polarized, first during G1 at the 'old' end opposite the site of septation, and then at the other end during G2 (Fig. 3a).

In fission yeast, a network of actin cables is oriented in the long axis of the cell, and in cortical patches at the cell 'ends'³⁶. Little is known about the machinery for localizing actin polymerization to cortical patches, although there are tantalizing indications of mechanisms similar to those in budding yeast (Fig. 3b). Cortical actin patches contain Orb2p, a member of the PAK family that may be downstream of Cdc42p³⁷, the actin-binding protein Bud6p³⁸, and

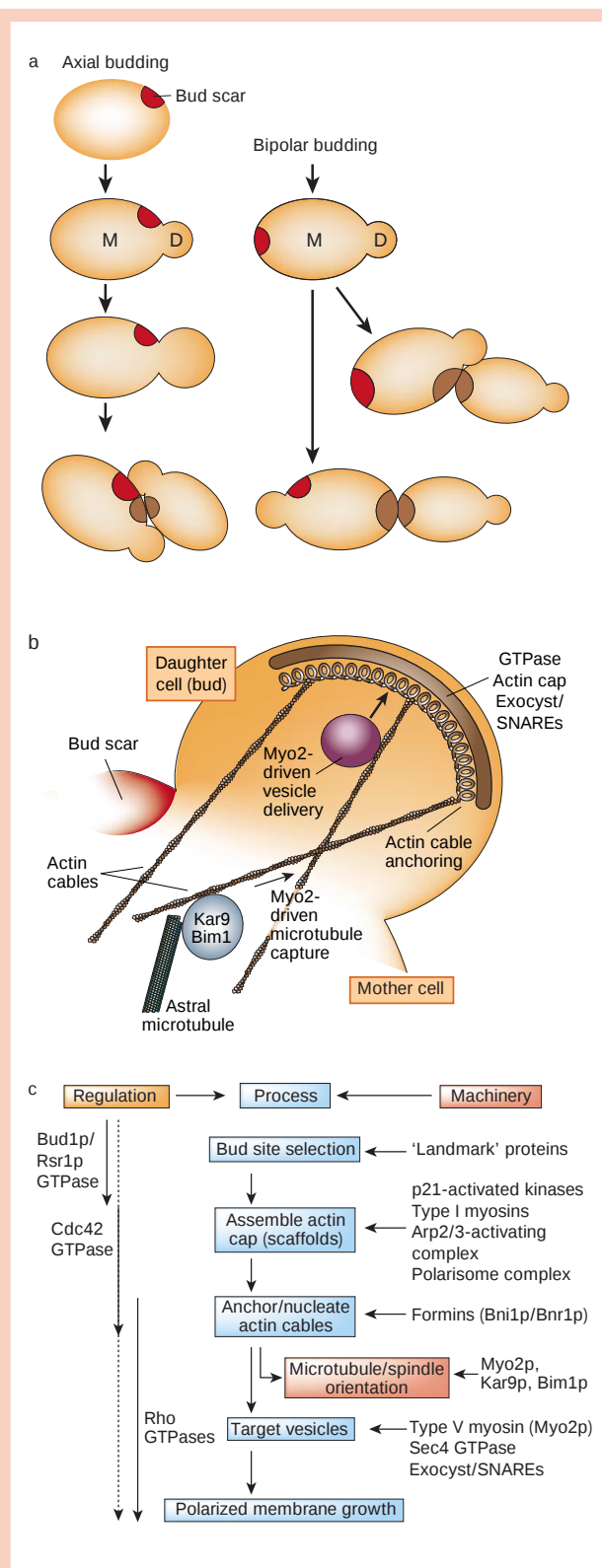


Figure 2 Protein pathways for generating cell polarity in budding yeast. **a**, Axial and bipolar budding patterns. **b**, A complex of proteins is assembled at the bud tip that orients the actin cytoskeleton, astral microtubules and vesicle delivery to the bud. **c**, Hierarchical organization of regulators, processes and cellular machinery linking bud-site selection to assembly of an actin cap, anchoring/nucleation of actin cables, and vesicle targeting that results in polarized membrane growth at the bud.

myosin-I and Arp2/3 complex³⁹ (Fig. 3b). For3p⁴⁰, a member of the formin family (like Bni1p in budding yeast; see above), localizes to cell 'ends' and cortical actin patches and is required for actin cable assembly and organization of actin patches. In *for3Δ* mutants, most daughter cells grow at both 'old' and 'new' ends simultaneously, indicating that, as in budding yeast, localization of a formin (For3p) at cell 'ends' could be part of a pathway that interprets and builds upon 'landmarks' for polarized membrane growth⁴¹.

Microtubules, like actin cables, are organized in the long axis of the cell and are uniformly polarized with their plus ends at the cell ends and minus ends around the nucleus. Disruption of microtubules causes cells to bend and branch abnormally as they grow³⁷, suggesting that they monitor the progress of cell growth. Microtubules have a 4–6 minute cycle of rapid polymerization towards a cell end, followed by catastrophic depolymerization ~100 seconds after contacting the end^{41,42}. Several proteins seem to control this cycle (Fig. 3b). Tip1p⁴² and Mal3p⁴³ bind to the plus ends of growing, but not shrinking microtubules, and may control microtubule length and dynamics.

How are polarized membrane growth and actin/microtubule cytoskeletons linked structurally and functionally? A candidate is the kelch repeat protein Tea1p, which localizes to microtubule plus ends and is deposited at cell ends when microtubules start to shrink⁴⁴. Tea1p is in a complex with Bud6p³⁸, which localizes to cell ends coincident with first 'old', and then 'new', end growth. Tea1p, Bud6p and perhaps For3p may provide a functional link between actin and microtubule cytoskeletons, and actin assembly and membrane growth at cell 'ends' (Fig. 3b). At present, little is known about how transport vesicle delivery is polarized for membrane growth in fission yeast, although the exocyst complex functions in vesicle delivery to cell ends and at the site of septation⁴⁵ (Fig. 3b).

Baseline principles for cell polarity

Core mechanisms for actin polymerization and vesicle transport are adapted for polarized membrane growth around a landmark at the plasma membrane. The landmark is recognized and reinforced by localized assembly of a signalling complex of small GTPases (for example, Bud1p/Rsr1p, Cdc42p and Rho), which relays signals to modular protein complexes that concentrate machinery for actin cytoskeleton assembly (Arp2/3, WASP/WIP, type-I myosins and profilin), anchoring/nucleation of actin cables (formin proteins, including Bni1p and Bnr1p/For3p, and Bud6p), microtubule assembly/capture (Kar9p/Bim1p and Tea1p/Tip1p/Mal3p) and vesicle delivery (type-V myosins, Sec4p, SNAREs and exocyst). Assembly of this hierarchy of complexes links landmarks to local assembly and orientation of the cytoskeleton, vesicle delivery for membrane growth, spindle orientation and inheritance of landmarks by daughter cells.

Generation of cell polarity in a multi-cell epithelium

So far, I have considered cell polarity in the context of single cells preparing to divide. Do the same principles of cell polarity that emerged from studies of single cells apply to polarized epithelial cells in multi-cell organisms? Clearly, there is increased complexity in multicellular tissues, as more than one membrane domain is formed (Fig. 4a), often concurrently, and the cell must use additional mechanisms to initially distinguish these domains, target and organize different proteins in each domain and, ultimately, keep the identities of the domains separate. I focus here on two control points. First, the apical junctional complex, a structure conserved between *Caenorhabditis elegans*, *Drosophila* and vertebrates (Fig. 4a), which initiates cell–cell adhesion and regulates the identities of different membrane domains. Second, mechanisms involved in sorting and targeting proteins to different membrane domains. As examples, I will examine formation of polarized epithelia during *de novo* assembly of plasma membrane domains in the *Drosophila* blastoderm and cultured mammalian epithelial cells.

Extracellular contacts between epithelial cells define the bounded (basolateral) and free (apical) cell surface⁴⁶. Although full develop-

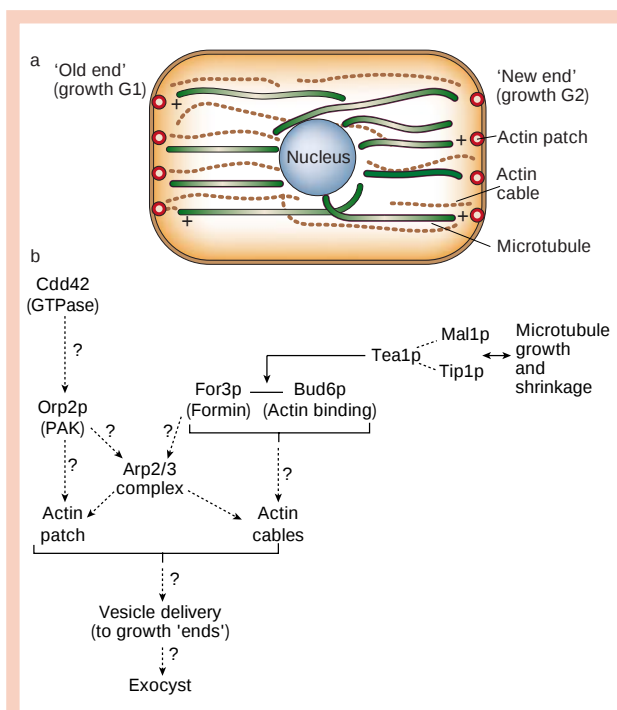


Figure 3 Protein pathways for generating cell polarity in fission yeast. **a**, Distribution of actin (cables and cortical patches) and microtubule cytoskeletons relative to the 'old' and 'new' ends of the cell, which grow during the G1 and G2 phases of the cell cycle, respectively. **b**, Hypothetical organization of proteins in cortical actin patches, and interactions of microtubule plus ends with the cell end/cortical actin patch (for details, see text).

ment of apicobasal polarity requires organization cues from both cell–cell and cell–extracellular matrix adhesions^{47,48}, assembly of the apical junctional complex, at the boundary between apical and (basal-)lateral membranes (Fig. 4a), recognizes and reinforces the principal landmark formed by cadherin-mediated cell–cell contacts. The apical junctional complex is a multifunctional, modular structure containing protein sub-complexes located at the boundary between the apical and lateral membrane domains⁴⁹ (Fig. 4a). In general, each protein sub-complex comprises an integral membrane protein bound to scaffold protein modules, each of which has multiple protein–protein binding motifs that potentially could inter-link different membrane protein sub-complexes. Scaffold proteins generally bind to the actin cytoskeleton^{50–52}, which is regulated by Rho-family small GTPases⁵³, although links to microtubules may also be present⁵⁴; some proteins at the apical junctional complex also act as transcriptional (co-)activators of gene expression⁵⁵.

Genetic studies in *Drosophila* and *C. elegans* have identified classes of proteins in the apical junctional complex that regulate development of epithelial cell polarity⁵⁶: cadherin/catenins^{57,58}, Crumbs (Crb)/Stardust⁵⁹, Bazooka (Baz/Par3)/Par6/atypical protein kinase C (aPKC)^{57,60,61} and Lethal giant larva (Lgl)⁶²/Scribble (Scrib)/Discs large (Dlg)^{64,65}/LET-413⁶⁶ (Fig. 4b). Recent studies show that the proteins in each class function in a common pathway, and the complexes seem to be integrated into a regulatory pathway that instructs apical and lateral membrane formation^{67,68}. Cell–cell adhesion mediated by the cadherin/catenin complex^{57,58,67,68} is required to initiate these events, as maternal and zygotic deletion of one of the catenins, *armadillo*, in *Drosophila* inhibits formation of all cellular structures including the first epithelium⁵⁸; note, however, the other complexes are required subsequently to maintain function and integrity of cadherin junctions^{67,68}.

The Baz complex is recruited to cadherin junctions^{57,67,68} and initiates formation of the apical membrane. Spread of apical membrane identity down the lateral membrane is impeded by activity of the Lgl/Scrib complex^{60,67,68}, which is recruited to a position below the Baz complex/cadherin junction^{60,67,68}. The Lgl/Scrib complex seems to antagonize functions of the Baz and, later, Crb complexes, thereby maintaining lateral membrane identity^{67,68} below the adherens junction. The Crb complex is recruited apically to the Baz complex/cadherin junction^{57,59} and seems to further antagonize activity of the Lgl/Scrib complex by blocking the spread of lateral membrane, thereby maintaining apical membrane identity^{67,68} above the adherens junction (Fig. 4b).

Although these studies identify overarching roles for these different complexes in initiating polarity of the apical and lateral membrane domains, the underlying biochemical mechanisms remain unclear. It is noteworthy that many of the proteins involved contain protein-binding motifs, PDZ domains (for example, Baz (Par3), Par6, Scrib, LET-413 and Discs large)⁵⁶ and leucine-rich domains (Scrib and LET-413)⁶⁸ that act as scaffolds to bind and spatially order proteins in the apical junctional complex^{69–71}. The Baz complex contains an aPKC^{60,72}. In mammalian epithelia, dominant negative mutants of aPKC cause mislocalization of Par3 (Baz) and ZO-1 (a tight-junction protein), resulting in disruption of tight and adherens junctions, and defects in apical/basolateral polarity⁷³. The Baz complex is also regulated by Rho-family GTPases that activate aPKC^{74,75}. In addition, Lgl binds to the t-SNARE syntaxin-4 and the complex localizes to the lateral membrane⁷⁶, but a role of Lgl in syntaxin-4 function has not been tested in detail. An attractive possibility is that complexes determining apical (Baz and Crb) and lateral (Lgl/Scrib) membrane domain identity may regulate delivery of transport vesicles containing apical and (basal-)lateral proteins, respectively, thereby contributing to the generation and maintenance of membrane domain identity.

Although specific protein complexes associated with the apical junctional complex control the identity of apical and basolateral membrane domains, how do cells determine the distribution of different proteins and functions in those domains? The generation of epithelial cell polarity during cellularization in the *Drosophila* embryo and in cultured mammalian cells provide mechanistic insights.

The first 13 cell divisions of the *Drosophila* embryo generate a syncytial blastoderm comprising ~6,000 nuclei that are localized beneath the plasma membrane. During cellularization, the surface area of the plasma membrane increases >25-fold as it forms between each nucleus to yield a sheet of columnar, polarized epithelial cells that are ~30 μm tall⁷⁷ (Fig. 5a). Cellularization occurs in two phases, 'slow' and 'fast'⁷⁸. In the slow phase, lasting 35–40 minutes and accounting for ~10 μm of lateral membrane, growth occurs by rapid invagination of the apical surface and redistribution of internalized protein to the lateral membrane (furrow canal)⁷⁸, perhaps by transcytosis. This phase of membrane growth is controlled by the product of the *slam* gene⁷⁹, which coordinates localized membrane insertion and membrane growth. The growing end of the furrow canal is marked by a small cluster of cadherin/catenin complexes that form a 'basal' junction that is organized by the product of the *nullo* gene⁸⁰; this junction also contains myosin-II and Discs lost⁸⁰. Loss of microtubules also disrupts cellularization⁸⁰. Thus, the force of membrane growth may be generated by insertion of membrane⁷⁸, with actin–myosin contraction⁸¹ and microtubules guiding and stabilizing the lateral membrane as it grows downwards⁷⁸.

Towards the end of the slow phase of growth, the depleted apical membrane is replenished directly from intracellular pools in the secretory pathway⁷⁸, and cadherin/catenin complexes begin to cluster towards a more apical aspect of the lateral membrane to form spot adherens junctions. During the subsequent fast phase of membrane growth, lasting 15–20 minutes and accounting for the remaining ~20 μm of membrane growth, new membrane is inserted

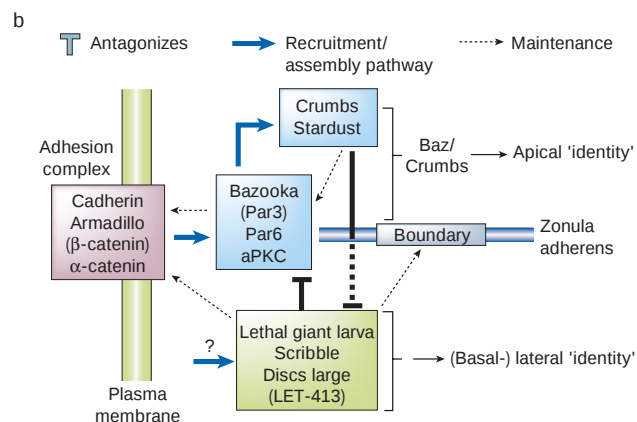
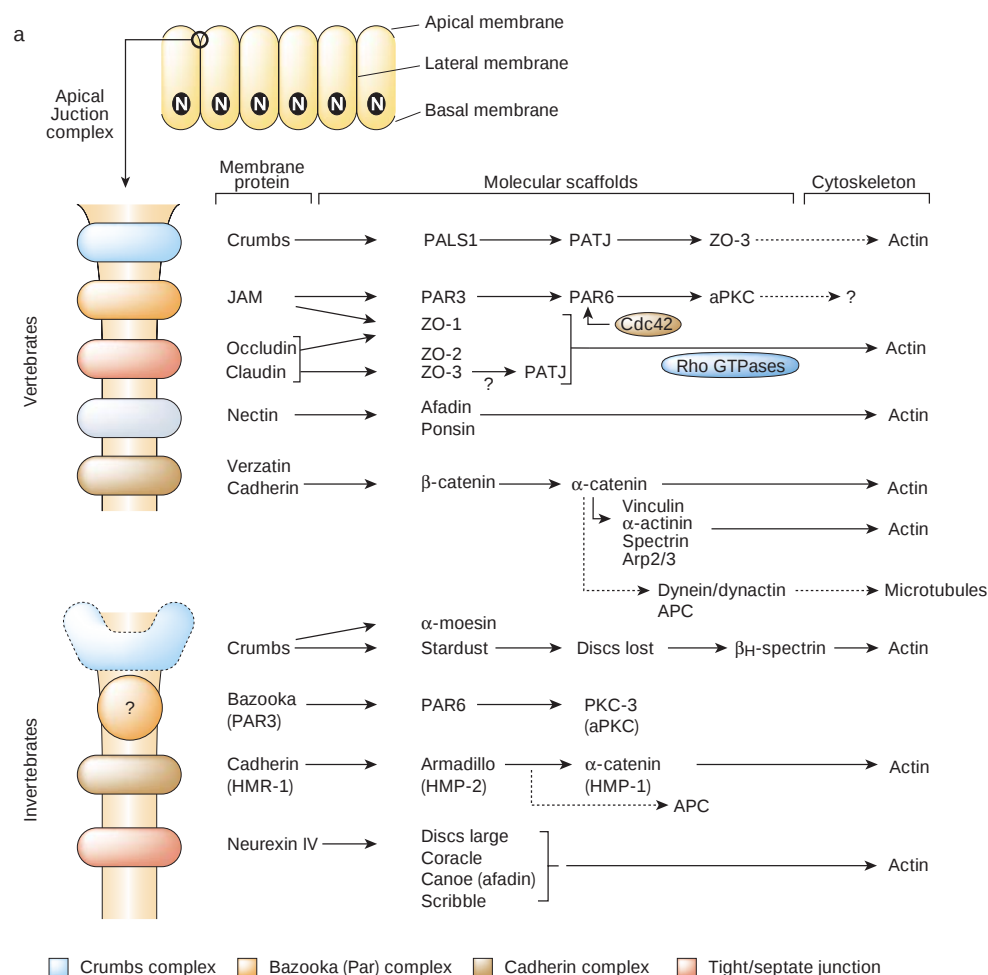


Figure 4 Organization of polarized epithelial cells and the apical junctional complex. **a**, Polarized epithelial cells form a monolayer in which the apical (unbounded surface) is separated at the boundary with the basal and lateral membranes (bounded surfaces) by the apical junctional complex (top). The main part of the panel shows molecular organization of the apical junctional complex. In vertebrates, the apical junctional complex is separated into structurally and functionally different sub-domains comprising membrane proteins (Crumbs, JAM (junctional adhesion molecule), nectin, occludin/claudin and cadherin) linked to modular protein scaffolds, which in turn bind mostly to the actin cytoskeleton, although links to microtubules are possible. In invertebrates (*C. elegans* and *Drosophila*), the apical junctional complex is similarly organized, except that the 'tight junction' function is provided by the septate junction localized below the cadherin (adherens) junction. **b**, Simplified scheme for how different protein complexes in the apical junctional complex regulate cell–cell adhesion (cadherin complex), and apical membrane (Bazooka and Crumbs complexes) and lateral membrane (Lethal giant larvae, Scribble and Disc large complex) identity. For details see text.

into the more apical portion of the lateral membrane by direct transport of vesicles from the secretory pathway^{56,78} (Fig. 5a).

How are vesicles targeted to designated sites on different membrane domains, and how are proteins pre-sorted in the Golgi complex prior to their delivery in vesicles to the plasma membrane? Studies of cell polarity in cultured mammalian epithelial cells have provided some answers to these questions, as they can be used for detailed biochemical and cell biological analyses of protein sorting pathways⁴⁶.

Most cells use endocytic pathways to route proteins between different domains of the plasma membrane and intracellular organelles⁸². In polarized epithelial cells, these pathways are coupled to sorting of proteins between different plasma membrane domains (termed transcytosis) (Fig. 5b). Transcytosis can be specific for transporting subsets of proteins between the apical and basolateral membranes (for example, ligand–receptor complexes⁸³) or general for many classes of proteins⁷⁸. In hepatocytes, for example, all membrane proteins are delivered from the Golgi to the basolateral membrane, and then apical proteins are specifically internalized and resorted by transcytosis to the apical membrane domain⁸⁴.

In many epithelial cells, protein distributions in different membrane domains are also determined by direct sorting of apical and basolateral proteins in the *trans*-Golgi network (TGN) and subsequent targeting to the correct membrane domain⁸⁵ (Fig. 5b). Sorting of proteins into the apical pathway seems to involve clustering membrane proteins into lipid rafts, and a glycosylphosphatidylinositol (GPI) domain in one class of proteins is sufficient to direct sorting into the apical pathway in most cells⁸⁶. Additional studies in budding yeast also suggest that lipid rafts are important in protein sorting⁸⁷. Sorting of proteins into the basolateral pathway requires sorting signals in the cytoplasmic domain and may be mediated by clustering through adaptor proteins and clathrin coats in the TGN or endosomes^{88,89}. Note that protein sorting in the TGN seems to be a constitutive process that occurs in cells such as fibroblasts that have neither a defined apical or basolateral membrane domain, nor cadherin-mediated cell–cell adhesion⁹⁰. Therefore, after cell adhesion, re-organization of the cytoskeleton and the generation of targeting patches for directed delivery of vesicles from the Golgi complex are critical from post-Golgi sorting of proteins to the correct membrane domain⁴⁶.

Organization of the cytoskeleton is important in regulating epithelial polarity (Fig. 5b). The actin cytoskeleton localizes to the cell cortex of each membrane domain, and actin re-organization occurs in a bundle circumscribing the cell at the apical junctional complex. This reorganization depends, in part, on linkage of actin to components of the apical junctional complex, and regulation of actin polymerization by Rho-family small GTPases⁵³, Arp2/3 and additional proteins⁹¹. Lateral membrane scaffolds containing actin-binding proteins regulate the distribution of some membrane proteins. For example, Na⁺,K⁺-ATPase is localized to the lateral membrane in mammalian and *Drosophila* epithelial cells through linkage to ankyrin and the spectrin scaffold^{92,93}. The spectrin scaffold is recruited to cell–cell contacts by binding to the actin cytoskeleton that links to (cadherin) sub-complexes of the apical junctional complex⁴⁶.

Microtubules are organized into bundles aligned in the apicobasal axis of the cell, with plus ends at the basal pole and intertwined mats of short filaments underneath the apical membrane and on the basal membrane⁴⁶ (Fig. 5b). How microtubules adopt these distributions is not understood. Microtubules may also be attached to the apical junctional complex⁵⁴, which may be important in orienting the mitotic spindle (this must be perpendicular to the apicobasal axis of polarity to maintain the monolayer organization of the epithelium). Actin may be important locally in vesicle delivery at the membrane, while microtubules seem to provide long-range pathways for vesicle delivery to plasma membrane domains, especially in transcytosis, and roles for both microtubule minus-end (dynein) and plus-end (kinesin family) motors has been implicated in vesicle targeting^{94,95}.

Specification of vesicle docking/fusion with different plasma membrane domains involves the SNARE complex, in which binding between cognate vesicle v-SNAREs and target membrane t-SNAREs specifies vesicle delivery. In polarized epithelial cells, different t-SNAREs are localized to apical (syntaxin-3) and basolateral (syntaxin-4) membranes⁹⁶ and inhibiting their function blocks vesicle delivery. Little is known about how t-SNAREs become localized to different membrane domains. Direct analysis of vesicle delivery to the basal and lateral membranes during polarization of Madin–Darby canine kidney (MDCK) cells revealed that, before cell–cell adhesion, vesicles containing either apical or basolateral membrane proteins fused with the basal surface⁹⁷. But after induction of cadherin-mediated cell–cell adhesion, neither apical nor basolateral vesicles fused with the basal surface; observations of vesicle fusion in the apical membrane was technically impossible, but basolateral vesicles were found to fuse at the top half of the lateral membrane⁹⁷. Thus, cell–cell adhesion may act as a signal for polarizing either vesicle docking/fusion machinery or vesicle/cytoskeleton transport pathways to the apical and lateral membranes (Fig. 5c). Interestingly, maintenance of syntaxin-3 depends on microtubules, and disruption of microtubules causes syntaxin-3 to mislocalize to lateral membrane, with the consequence that apical vesicles fuse with the lateral membrane; no effect on syntaxin-4 distribution was observed⁹⁷.

Like yeast, polarized epithelial cells express the exocyst complex (also referred to as the Sec6/8 complex)⁹⁸. In polarized MDCK cells, the exocyst is localized to the apical aspect of the lateral membrane⁹⁸, in the general region of basolateral vesicle docking/fusion⁹⁷. Inhibition of exocyst function inhibits delivery of vesicles containing a basolateral protein to the plasma membrane, but delivery of apical vesicles is unaffected⁹⁸. Cdc42 also has a role in delivery of basolateral, but not apical, transport vesicles⁹⁹. The polarized distribution of the exocyst complex depends on the maintenance of cell–cell contacts, and cadherin-mediated cell–cell adhesion recruits the exocyst specifically to sites of cell–cell contact⁹⁸. Similar to membrane growth during cellularization in the *Drosophila* embryo, polarization of MDCK cells results in a sixfold increase in the surface area of the lateral membrane, whereas those of apical and basal membrane domains do not increase¹⁰⁰ (Fig. 5c).

Conclusions

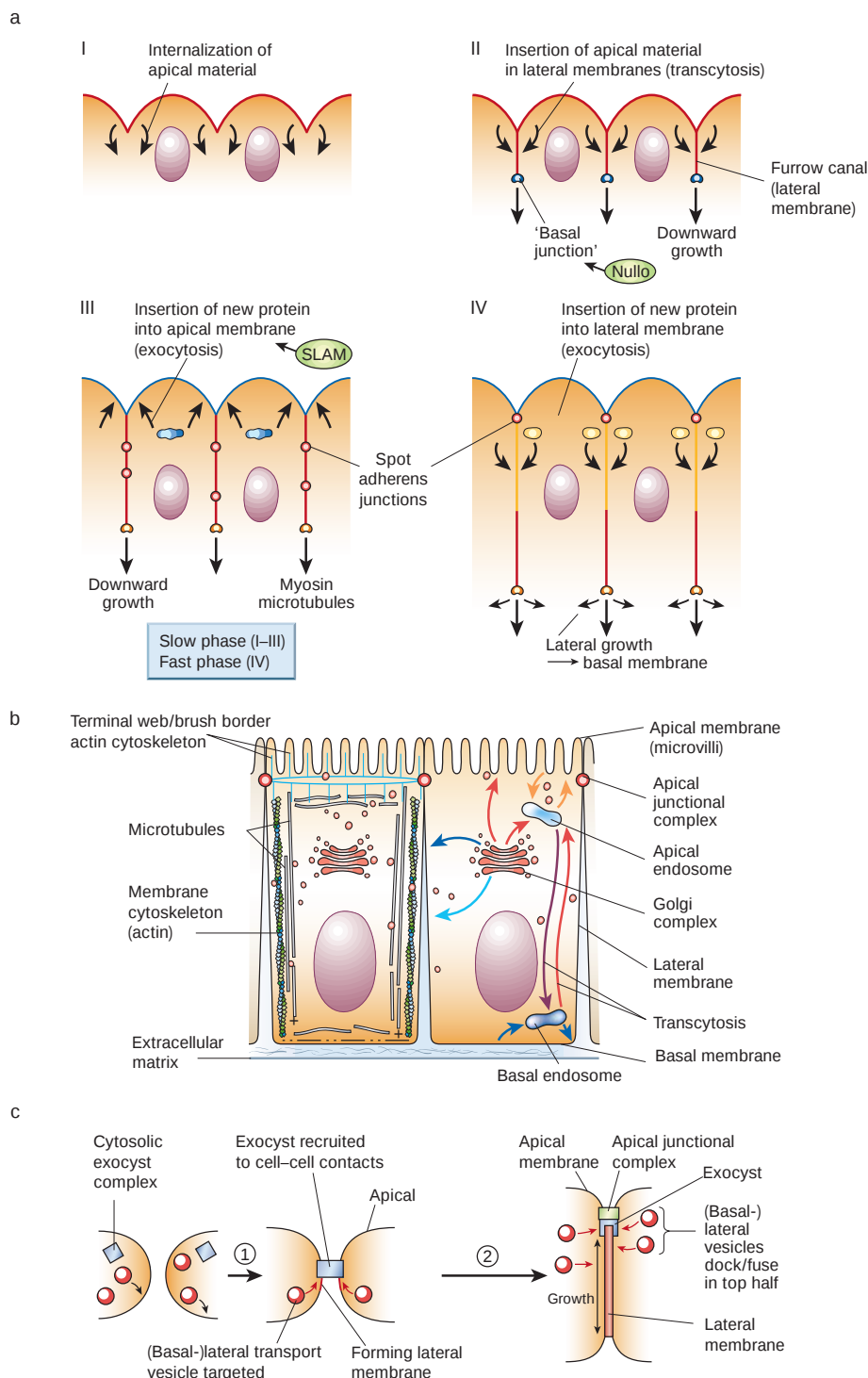
Cell polarity is portrayed in a wide variety of shapes and functions in single-cell organisms and cells in multi-cell tissues. Each cell type has not evolved a different mechanism to generate polarity, but has instead adapted a basic set of evolutionarily conserved core mechanisms, including: localized assembly of signalling complexes, cytoskeleton assembly and recruitment, mobilization of proteins from intracellular pools, and targeted vesicle delivery to sites of membrane growth. Polarized cells generate asymmetry around a cell-surface landmark by localized assembly of modular protein scaffolds that direct assembly and orientation of the cytoskeleton and specify vesicle delivery for membrane growth at that site.

Several principles of cell polarity can be deduced from these diverse examples. The pathway is hierarchical. It is initiated by a cell-surface landmark or spatial cue (the point of cytokinesis in budding and fission yeast, and cell–cell adhesion in epithelial cells from worms, flies and mammals), and defines a point on the cell surface to which the cell orients (mother–daughter axis in yeast, and the apicobasal axis in epithelial cells). This axis of polarity is propagated from the landmark throughout the cell by adaptation of core mechanisms that assemble and orient actin and microtubule cytoskeletons around the landmark.

Cytoskeleton reorganization around the landmark is determined by the local assembly at the landmark of signalling complexes of Rho-family small GTPases (Rho, Cdc42 and Rac1), which concentrate machinery for localized assembly of actin (profilin, Arp2/3 complex and type-I myosins), and modular protein scaffolds (the polarisome and Bni1p/Bud6p, and For3p/Bud6p in budding and fission yeast,

Figure 5 Generation of cell polarity in epithelia.

a. Formation of the cellular blastoderm in early *Drosophila* embryogenesis (for details, see text). **b.** Schematic representation of polarized epithelial cells. Left, organization of the actin and microtubule cytoskeletons; right, organization of vesicle transport pathways to different plasma membrane domains either directly from the Golgi complex, or indirectly via apical or basal endosomes through endocytic or transcytotic pathways. **c.** Generation of the lateral membrane domain in cultured epithelial cells. Prior to cadherin-mediated cell–cell adhesion, the exocyst is cytosolic and vesicles fuse with the basal membrane. Upon cell–cell adhesion (step 1), the exocyst is recruited to cell–cell contacts and vesicles fuse with the forming lateral membrane. As more vesicles fuse, the lateral membrane increases in area around sixfold. Later (step 2), the exocyst and vesicle delivery are located in the apical region of the lateral membrane.



respectively, and the apical junctional complex comprising Crb, Baz and Lgl/Scrib complexes in epithelia). Actin polymerization and assembly of protein scaffolds also reinforce the organization, retention and inheritance of the landmark. Actin cytoskeleton orientation around the landmark may regulate capture/reorientation of microtubules (involving complexes of Kar9p, Tip1p/CLIP170, Bim1p/Mal1p/EB1 and APC) for spindle orientation in the axis of cell polarity (in budding yeast, *Drosophila* neuroblasts and epithelia)

and for vesicle delivery to different membrane domains (in transcytotic and exocytic vesicle transport in epithelia).

The efficiency and fidelity of vesicle delivery to the landmark depends on cytoskeleton assembly and orientation, and recruitment to the landmark of vesicle docking/fusion machinery (small GTPases such as Sec4p and RalA, SNARE protein complexes and the exocyst); in epithelia, the machinery is separated to the apical and lateral membrane domains. Localized vesicle delivery results in rapid plasma

membrane growth around the landmark. The identity of the new membrane domain is maintained by the subset of (pre-sorted) proteins that had been delivered there in vesicles either directly from the Golgi complex or by transcytosis, and by retention through membrane scaffolds. Assembly of this hierarchy of complexes links landmarks to local assembly and orientation of the cytoskeleton, vesicle delivery for membrane growth, spindle orientation, and the inheritance of landmarks by daughter cells. □

doi:10.1038/nature01602

1. Chant, J. Cell polarity in yeast. *Annu. Rev. Cell Dev. Biol.* **15**, 365–391 (1999).
2. Chant, J. & Herskowitz, I. Genetic control of bud site selection in yeast by a set of gene products that constitute a morphogenetic pathway. *Cell* **65**, 1203–1212 (1991).
3. Zahner, J. E., Harkins, H. A. & Pringle, J. R. Genetic analysis of the bipolar pattern of bud site selection in the yeast *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **16**, 1857–1870 (1996).
4. Bender, A. & Pringle, J. R. Multicopy suppression of the *cdc24* budding defect in yeast by *CDC42* and three newly identified genes including the ras-related gene *RSR1*. *Proc. Natl Acad. Sci. USA* **89**, 9976–9980 (1992).
5. Pruyne, D. & Bretscher, A. Polarization of cell growth in yeast I. Establishment and maintenance of polarity states. *J. Cell Sci.* **113**, 365–375 (2000).
6. Adams, A., Johnson, D., Longnecker, R., Sloat, B. & Pringle, J. *CDC42* and *CDC43*, two additional genes involved in budding and the establishment of cell polarity in the yeast *Saccharomyces cerevisiae*. *J. Cell Biol.* **111**, 131–142 (1990).
7. Etienne-Manneville, S. & Hall, A. Rho GTPases in cell biology. *Nature* **420**, 629–635 (2003).
8. Kang, P. J., Sanson, A., Lee, B. & Park, H.-O. A GDP/GTP exchange factor involved in linking a spatial landmark to cell polarity. *Science* **292**, 1376–1378 (2001).
9. Marston, A. L., Chen, T., Yang, M. C., Belhumeur, P. & Chant, J. A localized GTPase exchange factor, Bud5, determines the orientation of division axes in yeast. *Curr. Biol.* **11**, 803–807 (2001).
10. Park, H.-O., Sanson, A. & Herskowitz, J. Localization of Bud2p, a GTPase-activating protein necessary for programming cell polarity in yeast to the presumptive bud site. *Genes Dev.* **13**, 1912–1917 (1999).
11. Zheng, Y., Bender, A. & Cerione, R. Interactions among proteins involved in bud-site selection and bud-site assembly in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **270**, 626–630 (1995).
12. Wedlich-Soldner, R., Altschuler, S., Wu, L. & Li, R. Spontaneous cell polarization through actomyosin-based delivery of the Cdc42 GTPase. *Science* **299**, 1231–1235 (2003).
13. Pruyne, D. & Bretscher, A. Polarization of cell growth in yeast II. The role of the actin cytoskeleton. *J. Cell Sci.* **113**, 571–585 (2000).
14. Eby, J. et al. Actin cytoskeleton organization regulated by the PAK family of protein kinases. *Curr. Biol.* **8**, 967–970 (1998).
15. Wu, C., Lyttyn, V., Thomas, D. & Leberer, E. The phosphorylation site for Ste20p-like protein kinase is essential for the function of myosin-I in yeast. *J. Biol. Chem.* **272**, 30623–30626 (1997).
16. Evangelista, M. et al. A role for myosin-I in actin assembly through interactions with Vrp1p, Bee1p, and the Arp2/3 complex. *J. Cell Biol.* **148**, 353–362 (2000).
17. Lee, W. L., Bezanilli, M. & Pollard, T. D. Fission yeast myosin-I, Myo1p, stimulates actin assembly by Arp2/3 complex and shares functions with WASP. *J. Cell Biol.* **151**, 789–800 (2000).
18. Lechler, T., Jonsdottir, G. A., Klee, S. K., Pellman, D. & Li, R. A two-tiered mechanism by which Cdc42 controls the localization and activation of an Arp2/3-activating motor complex in yeast. *J. Cell Biol.* **155**, 261–270 (2001).
19. Sheu, Y. P., Santos, B., Fortin, N., Costigan, C. & Snyder, M. Spa2p interacts with cell polarity proteins and signaling components involved in yeast cell morphogenesis. *Mol. Cell Biol.* **18**, 4035–4069 (1998).
20. Evangelista, M., Eby, J. et al. A yeast formin linking Cdc42p and the actin cytoskeleton during polarized morphogenesis. *Science* **276**, 118–121 (1997).
21. Sagot, I., Klee, S. K. & Pellman, D. Yeast formins regulate cell polarity by controlling the assembly of actin cables. *Nature Cell Biol.* **4**, 42–50 (2002).
22. Evangelista, M., Pruyne, D., Amberg, D. C., Boone, C. & Bretscher, A. Formins direct Arp2/3-independent actin filament assembly to polarize cell growth in yeast. *Nature Cell Biol.* **4**, 32–41 (2002).
23. Imamura, H. et al. Bni1p and Bnr1p: downstream targets of the Rho family of small GTPases which interact with profilin and regulate actin cytoskeleton in *Saccharomyces cerevisiae*. *EMBO J.* **16**, 2745–2755 (1997).
24. Amberg, D. C., Zahner, J. E., Mulholland, J. W., Pringle, J. R. & Botstein, D. Aip3p/Bud6p, a yeast actin-binding protein that is involved in morphogenesis and the selection of bipolar budding sites. *Mol. Biol. Cell* **8**, 729–753 (1997).
25. Kohno, H. et al. Bni1p implicated in cytoskeletal control is a putative target of Rho1p small GTP binding protein in *Saccharomyces cerevisiae*. *EMBO J.* **15**, 6060–6068 (1996).
26. Sagot, I., Rodal, A. A., Moseley, J., Goode, B. L. & Pellman, D. An actin nucleation mechanism mediated by Bni1 and profilin. *Nature Cell Biol.* **4**, 626–631 (2002).
27. Pruyne, D. et al. Role of formins in actin assembly: nucleation and barbed-end association. *Science* **297**, 612–615 (2002).
28. Carminati, J. L. & Stearns, T. Microtubules orient the mitotic spindle in yeast through dynein-dependent interactions with the cell cortex. *J. Cell Biol.* **138**, 629–641 (1997).
29. Liakopoulos, D., Kusch, J., Grava, S., Vogel, J. & Barral, Y. Asymmetric loading of Kar9 onto spindle poles and microtubules ensures proper spindle alignment. *Cell* **112**, 561–574 (2003).
30. Brennwald, P. et al. Sec9 is a SNAP-25-like component of a yeast SNARE complex that may be the effector of Sec4 function in exocytosis. *Cell* **79**, 245–258 (1994).
31. Schott, D., Ho, J., Pruyne, D. & Bretscher, A. The COOH-terminal domain of Myo2p, a yeast myosin V, has a direct role in secretory vesicle targeting. *J. Cell Biol.* **147**, 791–808 (1999).
32. Karpova, T. S. et al. Role of actin and Myo2p in polarized secretion and growth of *Saccharomyces cerevisiae*. *Mol. Biol. Cell* **11**, 1727–1737 (2000).
33. TerBush, D. R., Maurice, T., Roth, D. & Novick, P. The exocyst is a multi-protein complex required for exocytosis in *Saccharomyces cerevisiae*. *EMBO J.* **15**, 6483–6494 (1996).
34. Novick, P. & Guo, W. Ras family therapy: Rab, Rho and Ral talk to the exocyst. *Trends Cell Biol.* **12**, 247–249 (2002).
35. Lehman, K., Rossi, G., Adamo, J. E. & Brennwald, P. Yeast homologues of tomosyn and lethal giant larvae function in exocytosis and are associated with the plasma membrane SNARE, Sec9. *J. Cell Biol.* **146**, 125–140 (1999).
36. Chang, F. Establishment of a cellular axis in fission yeast. *Trends Genet.* **17**, 273–278 (2001).
37. Sawin, K. E., Hajibaghi, M. A. & Nurse, P. Mis-specification of cortical identity in a fission yeast PAK mutant. *Curr. Biol.* **9**, 1335–1338 (1999).
38. Glynn, J., Lustig, R., Berlin, A. & Chang, F. Role of bud6p and tea1p in the interaction between actin and microtubules for the establishment of cell polarity in fission yeast. *Curr. Biol.* **11**, 836–845 (2001).
39. Pelham, R. J. & Chang, F. Role of actin polymerization and actin cables in the movement of actin patches in *S. pombe*. *Nature Cell Biol.* **3**, 235–244 (2001).
40. Feierbach, B. & Chang, F. Role of the fission yeast formin for3p in cell polarity, actin cable formation, and symmetric cell division. *Curr. Biol.* **11**, 1656–1665 (2001).
41. Tran, P. T., Marsh, L., Doyle, V., Inoue, S. & Chang, F. A mechanism for nuclear positioning in fission yeast based on microtubule pushing. *J. Cell Biol.* **153**, 397–411 (2001).
42. Brunner, D. & Nurse, P. CLIP170-like Tip1p spatially organizes microtubular dynamics in fission yeast. *Cell* **102**, 695–704 (2000).
43. Beinhaber, J. D., Hagan, I. M., Hegeman, J. H. & Feig, U. Mal3, the fission yeast homologue of the human APC-interacting protein EB1 is required for microtubule integrity and the maintenance of cell form. *J. Cell Biol.* **139**, 717–728 (1997).
44. Mata, J. & Nurse, P. tea1 and the microtubular cytoskeleton are important for generating global spatial order within the fission yeast cell. *Cell* **89**, 939–949 (1997).
45. Wang, H. et al. The multiprotein exocyst complex is essential for cell separation in *Schizosaccharomyces pombe*. *Mol. Biol. Cell* **13**, 515–529 (2002).
46. Yeaman, C., Grindstaff, K. K. & Nelson, W. J. New perspectives on mechanisms involved in generating epithelial cell polarity. *Physiol. Rev.* **79**, 73–98 (1999).
47. Wang, A. Z., Ozkan, G. K. & Nelson, W. J. Steps in the morphogenesis of a polarized epithelium. I. Uncoupling the roles of cell-cell and cell-substratum contact in establishing plasma membrane polarity in multicellular epithelial (MDCK) cysts. *J. Cell Sci.* **95**, 137–151 (1990).
48. O'Brien, L. E., Zegers, M. M. & Mostov, K. E. Building epithelial architecture: insights from three-dimensional culture models. *Nature Rev. Mol. Cell Biol.* **3**, 531–537 (2002).
49. Knust, E. & Bossinger, O. Epithelial polarity: composition and formation of intercellular junctions in different organisms. *Science* **298**, 1955–1959 (2003).
50. Dimitratos, S. D., Woods, D. F., Stathakis, D. G. & Bryant, P. J. Signaling pathways are focused at specialized regions of the plasma membrane by scaffolding proteins of the MAGUK family. *BioEssays* **21**, 912–921 (1999).
51. Mohler, P. J., Gramolini, A. O. & Bennett, V. Ankyrins. *J. Cell Sci.* **115**, 1565–1566 (2002).
52. Tsukita, S., Furuse, M. & Itoh, M. Structural and signaling molecules come together at tight junctions. *Curr. Opin. Cell Biol.* **11**, 628–633 (1999).
53. Fukata, M. & Kaibuchi, K. Rho-family GTPases in cadherin-mediated cell–cell adhesion. *Nature Rev. Mol. Cell Biol.* **2**, 887–897 (2001).
54. Ligon, L. A., Karki, S., Tokito, M. & Holzbaur, E. L. Dynein binds to β -catenin and may tether microtubules at adherens junctions. *Nature Cell Biol.* **3**, 913–917 (2001).
55. Balda, M. S., Garrett, M. D. & Matter, K. The ZO-1-associated Y-box factor ZONAB regulates epithelial cell proliferation and cell density. *J. Cell Biol.* **160**, 423–432 (2003).
56. Tepass, U., Tanentzapf, G., Ward, R. & Fehon, R. Epithelial cell polarity and cell junctions in *Drosophila*. *Annu. Rev. Genet.* **35**, 747–784 (2001).
57. Muller, H. A. J. & Wieschaus, E. Armadillo, bazooka, and stardust are critical for early stages in formation of the zonula adherens and maintenance of the polarized blastoderm epithelium in *Drosophila*. *J. Cell Biol.* **134**, 149–163 (1996).
58. Cox, R. T., Kirkpatrick, C. & Peifer, M. Armadillo is required for adherens junction assembly, cell polarity, and morphogenesis during *Drosophila* embryogenesis. *J. Cell Biol.* **134**, 133–148 (1996).
59. Tepass, U. & Knust, E. *crumbs* and *stardust* act in a genetic pathway that controls the organization of epithelia in *Drosophila melanogaster*. *Dev. Biol.* **159**, 311–326 (1993).
60. Wodarz, A., Ramath, A., Grimm, A. & Knust, E. *Drosophila* atypical protein kinase C associates with Bazooka and controls polarity of epithelia and neuroblasts. *J. Cell Biol.* **150**, 1361–1374 (2000).
61. Petronczki, M. & Knoblich, J. DmPAR-6 directs epithelial polarity and asymmetric cell division of neuroblasts in *Drosophila*. *Nature Cell Biol.* **3**, 43–49 (2001).
62. Bilder, D. & Perrimon, N. Cooperative regulation of cell polarity and growth by *Drosophila* tumor suppressors. *Science* **289**, 113–116 (2000).
63. Bilder, D., Li, M. & Perrimon, N. Localization of apical determinants by the basolateral PDZ protein Scribble. *Nature* **403**, 676–680 (2000).
64. Woods, D. F. & Bryant, P. J. The *discs-large* tumor suppressor gene of *Drosophila* encodes a guanylate kinase homolog localized to septate junctions. *Cell* **66**, 451–464 (1991).
65. Bossinger, O., Klebes, A., Segbert, C., Theres, C. & Knust, E. Zonula adherens formation in *Caenorhabditis elegans* requires *dgl-1*, the homologue of the *Drosophila* gene *discs large*. *Dev. Biol.* **230**, 29–42 (2001).
66. Legouis, R. et al. LET-413 is a basolateral protein required for the assembly of adherens junctions in *Caenorhabditis elegans*. *Nature Cell Biol.* **2**, 415–422 (2000).
67. Tanentzapf, G. & Tepass, U. Interactions between the *crumbs*, *lethal giant larvae* and *bazooka* pathways in epithelial polarization. *Nature Cell Biol.* **5**, 46–52 (2003).
68. Bilder, D., Schuber, M. & Perrimon, N. Integrating activity of PDZ protein complexes regulates epithelial polarity. *Nature Cell Biol.* **5**, 53–58 (2003).
69. Hurd, T. W., Gao, L., Roh, M. H., Macara, I. G. & Margolis, B. Direct interaction of two polarity complexes implicated in epithelial tight junction assembly. *Nature Cell Biol.* **5**, 137–142 (2003).
70. Medina, E. et al. *Crumbs* interacts with moesin and β -tubulin-spectrin in the apical membrane skeleton of *Drosophila*. *J. Cell Biol.* **158**, 941–951 (2002).
71. Gao, L., Joberty, G. & Macara, I. G. Assembly of epithelial tight junctions is negatively regulated by Par6. *Curr. Biol.* **12**, 221–225 (2002).
72. Izumi, Y. et al. An atypical PKC directly associates and colocalizes at the epithelial junction with ASIP, a mammalian homologue of the *Caenorhabditis elegans* polarity protein PAR-3. *J. Cell Biol.* **143**, 95–103 (1998).
73. Suzuki, A. et al. aPKC kinase activity is required for the asymmetric differentiation of the premature junctional complex during epithelial cell polarization. *J. Cell Sci.* **115**, 3565–3573 (2002).
74. Joberty, G., Petersen, C., Gao, L. & Macara, I. G. A cell-polarity protein Par6 links Par3 and atypical protein kinase C to Cdc42. *Nature Cell Biol.* **2**, 531–539 (2000).

75. Lin, D. *et al.* A mammalian PAR-3–PAR-6 complex implicated in Cdc42/Rac1 and aPKC signaling and cell polarity. *Nature Cell Biol.* **2**, 540–547 (2000).
76. Müsch, A. *et al.* A mammalian homologue of the *Drosophila* tumor suppressor lethal (2) giant larvae interacts with basolateral exocytic machinery in MDCK cells. *Mol. Biol. Cell.* **13**, 158–168 (2002).
77. Foe, V. E., Odell, G. M. & Edgar, B. A. in *The Development of Drosophila melanogaster* Vol. 1 (ed. Bate, M.) 149–300 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1993).
78. Lecuit, T. & Wieschaus, E. Polarized insertion of new membrane from a cytoplasmic reservoir during cleavage of the *Drosophila* embryo. *J. Cell Biol.* **150**, 849–860 (2000).
79. Lecuit, T., Samata, R. & Wieschaus, E. *slam* encodes a developmental regulator of polarized membrane growth during cleavage of the *Drosophila* embryo. *Dev. Cell* **2**, 425–436 (2002).
80. Hunter, C. & Wieschaus, E. Regulated expression of *nullo* is required for the formation of distinct apical and basal adherens junctions in the *Drosophila* blastoderm. *J. Cell Biol.* **150**, 391–401 (2000).
81. Schejter, E. D. & Wieschaus, E. Bottleneck acts as a regulator of the microfilament network governing cellularization of the *Drosophila* embryo. *Cell* **75**, 373–385 (1993).
82. Pfeffer, S. Membrane domains in the secretory and endocytic pathways. *Cell* **112**, 507–517 (2003).
83. Mostov, K. E. & Deitcher, D. L. Polymeric immunoglobulin receptor expressed in MDCK cells transcytoses IgA. *Cell* **46**, 613–621 (1986).
84. Bartles, J. R., Feracci, H. M., Stieger, B. & Hubbard, A. L. Biogenesis of the rat hepatocyte plasma membrane in vivo: comparison of the pathways taken by apical and basolateral proteins using subcellular fractionation. *J. Cell Biol.* **105**, 1241–1251 (1987).
85. Mostov, K. E., Verges, M. & Altschuler, Y. Membrane traffic in polarized epithelial cells. *Curr. Opin. Cell Biol.* **12**, 483–490 (2000).
86. Lisanti, M. P., Caras, I. W., Davitz, M. A. & Rodriguez-Boulán, E. A glycosphingolipid membrane anchor acts as an apical targeting signal in polarized epithelial cell. *J. Cell Biol.* **109**, 2145–2156 (1989).
87. Bagnat, M., Chang, A. & Simons, K. Plasma membrane proton ATPase Pma1p requires raft association for surface delivery in yeast. *Mol. Biol. Cell* **12**, 4129–4138 (2001).
88. Hunziker, W., Harter, C., Matter, K. & Mellman, I. Basolateral sorting in MDCK cells requires a distinct cytoplasmic domain determinant. *Cell* **66**, 907–920 (1991).
89. Areoti, B., Kosen, P. A., Kuntz, I. D., Cohen, F. E. & Mostov, K. E. Mutational and secondary structural analysis of the basolateral sorting signal of the polymeric immunoglobulin receptor. *J. Cell Biol.* **123**, 1149–1160 (1993).
90. Yoshimori, T., Keller, P., Roth, M. G. & Simons, K. Different biosynthetic transport routes to the plasma membrane in BHK and CHO cells. *J. Cell Biol.* **133**, 247–256 (1996).
91. Perez-Moreno, M., Jamora, C. & Fuchs, E. Sticky business: orchestrating cellular signals at adherens junctions. *Cell* **112**, 535–548 (2003).
92. Nelson, W. J. & Veshnock, P. J. Ankyrin (membrane-skeleton) binds to the Na⁺, K⁺-ATPase: implications for the organization of membrane domains in polarized cells. *Nature* **328**, 533–536 (1987).
93. Dubreuil, R. R., Wang, P., Dahl, S., Lee, J. & Goldstein, L. S. *Drosophila* β spectrin functions independently of α spectrin to polarize the Na,K ATPase in epithelial cells. *J. Cell Biol.* **149**, 647–656 (2000).
94. Lafont, F., Burkhardt, J. K. & Simons, K. Involvement of microtubule motors in basolateral and apical transport in kidney cells. *Nature* **372**, 801–803 (1994).
95. Kreitzer, G., Marmostein, A., Okamoto, P., Vallee, R. & Rodriguez-Boulán, E. Kinesin and dynamin are required for post-Golgi transport of a plasma-membrane protein. *Nature Cell Biol.* **2**, 125–127 (2000).
96. Low, S. H. *et al.* Differential localization of syntaxin isoforms in polarized Madin-Darby canine kidney cells. *Mol. Biol. Cell* **7**, 20007–2018 (1996).
97. Kreitzer, G. *et al.* Three-dimensional analysis of post-Golgi carrier exocytosis in epithelial cells. *Nature Cell Biol.* **5**, 126–136 (2003).
98. Grindstaff, K. K. *et al.* Sec6/8 complex is recruited to cell-cell contacts and specifies transport vesicle delivery to the basal-lateral membrane in polarized epithelial cells. *Cell* **93**, 731–740 (1998).
99. Cohen, D., Musch, A. & Rodriguez-Boulán, E. Selective control of basolateral membrane proteins polarity by Cdc42. *Traffic* **2**, 556–564 (2001).
100. Vega-Salas, D. E., Salas, P. J., Gundersen, D. & Rodriguez-Boulán, E. Formation of the apical pole of epithelial (Madin-Darby canine kidney) cells: polarity of an apical protein is independent of tight junctions while segregation of a basolateral marker requires cell-cell interactions. *J. Cell Biol.* **104**, 905–916 (1987).

Acknowledgements This review is dedicated to I. Herskowitz (University of California, San Francisco) who first inspired me to think broadly about cell polarity.

Microbial pathogenesis and cytoskeletal function

Samantha Gruenheid & B. Brett Finlay

Biotechnology Laboratory, University of British Columbia, Vancouver, Canada V6T 1Z3 (e-mail: bfinlay@interchange.ubc.ca)

Pathogenic microbes subvert normal host-cell processes to create a specialized niche, which enhances their survival. A common and recurring target of pathogens is the host cell's cytoskeleton, which is utilized by these microbes for purposes that include attachment, entry into cells, movement within and between cells, vacuole formation and remodelling, and avoidance of phagocytosis. Our increased understanding of these processes in recent years has not only contributed to a greater comprehension of the molecular causes of infectious diseases, but has also revealed fundamental insights into normal functions of the cytoskeleton. From the use of bacterial toxins to investigate Rho family GTPases to *in vitro* studies of actin polymerization using *Listeria* and *Shigella*, the study of pathogenesis has provided important tools to probe cytoskeletal function.

To induce cytoskeletal changes, pathogenic microbes must ensure delivery of effector molecules onto or into host cells. Effectors are usually proteins that interface with and influence host-cell pathways, and can facilitate disease. As viruses have no metabolic activity outside their hosts, effectors must already be packaged in the virion or expressed from the viral genome from within host cells. Bacteria use several methods to deliver effector proteins to the host cell. Some effectors, such as toxins, are secreted by bacteria in the vicinity of the host cell, where they bind specific receptors and are taken up by endocytosis¹. Other effector proteins can facilitate their own uptake with pore-forming subunits or autotransporter domains. Some Gram-negative pathogenic bacteria have acquired sophisticated 'molecular syringes', such as type III or type IV secretion systems, which are multisubunit molecular machines that span the bacterial and host membranes and translocate effectors directly into host cells. Protozoan parasites use secretory organelles, such as the micronemes, rhoptries and dense granules of *Plasmodium* and *Toxoplasma*, to deliver microbial products to the host-pathogen interface.

Forced entry

Viruses and some bacteria and protozoan pathogens are obligate intracellular parasites that can only replicate inside their host cells. Other pathogens can replicate extracellularly, but choose an intracellular lifestyle to obtain a favourable niche within the host. To gain access into non-phagocytic cells, or to enter into a protected niche within phagocytic cells, microbes have developed dedicated strategies that mediate pathogen invasion. Because the cytoskeleton controls surface remodelling events such as phagocytosis² and macropinocytosis³, it is an obvious target for such invasion mechanisms. Many Gram-negative bacteria utilize a type III secretion system (TTSS) and associated effectors to mediate invasion into non-phagocytic cells⁴. *Salmonella* and *Shigella* species direct their own uptake into host cells using a multifaceted approach, coordinating several signalling pathways that converge to induce transient, actin-rich membrane ruffles that engulf the infecting bacteria (Fig. 1a, b).

Salmonella directly activate Rho GTPases using secreted effectors and a TTSS encoded within the *Salmonella* patho-

genicity island 1 (SPI-1) locus. The SPI-1-secreted effectors SopE and SopE2 act as guanine-nucleotide-exchange factors (GEFs) for the small GTPases Cdc42 and Rac⁵. Structural analysis reveals that, despite a lack of sequence and architectural similarity, SopE and eukaryotic GEFs induce virtually identical conformational changes in their target Rho proteins, providing an example of bacterial mimicry of a normal cellular process through convergent evolution⁶. Additional SPI-1-translocated effectors of *Salmonella* affect actin dynamics during the invasion process. SipA binds to and stabilizes actin, and SipC, which forms part of the TTSS delivery pore, nucleates and bundles actin while anchored in the host cell membrane⁵.

Salmonella also alters the actin cytoskeleton through manipulation of phosphoinositides. The plasma membrane is intimately associated with the actin cytoskeleton, and this interaction depends on phosphatidylinositol 4,5-bisphosphate (PtdIns(4,5)P₂)⁷. SigD/SopB is an SPI-1-translocated inositol phosphatase that induces the rapid disappearance of PtdIns(4,5)P₂ from invaginating regions of the membrane during *Salmonella* invasion. This increases elasticity to facilitate the remodelling of the plasma membrane associated with *Salmonella* entry⁸. PtdIns(4,5)P₂ has also been implicated in vesicle fission during the creation of phagosomes and clathrin-coated vesicles, and accordingly, SigD also is involved in sealing plasma membrane invaginations to form bona fide vacuoles⁸. After invasion, an additional SPI-1 effector, SptP, acts as a GTPase-activating protein (GAP) for Cdc42 and Rac1, thereby inactivating these G proteins and returning cell morphology to a relatively normal state⁹.

SptP is a bifunctional protein, with its GAP domain at the amino terminus, and a protein tyrosine phosphatase domain at the carboxy terminus⁵. A potential target for the tyrosine phosphatase activity of SptP is the intermediate filament protein vimentin, which is recruited to the membrane ruffles stimulated by *Salmonella*¹⁰. Other studies have also identified another intermediate filament protein involved in *Salmonella* entry: SipC binds cytokeratins and expression of dominant-negative cytokeratin-18 inhibits *Salmonella* entry into HEP2 cells¹¹. As virtually nothing is known about the disruption of intermediate filaments by pathogens, this emerging area of investigation shows promise for future advances.

In a mechanism very similar to *Salmonella*, *Shigella* uses a TTSS to deliver effectors that activate Cdc42 and Rac¹² and

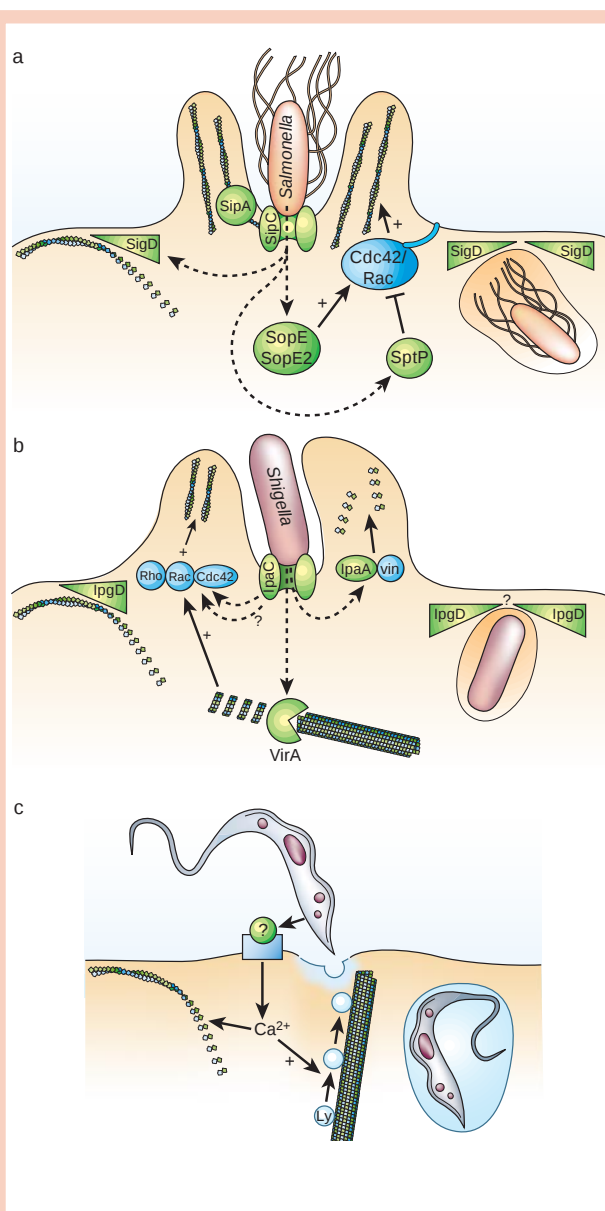


Figure 1 Host-cell invasion. **a**, *Salmonella* invasion. Entry into host cells is mediated by the *Salmonella* pathogenicity island-1 (SPI-1) type III secretion system (TTSS) and its effectors. Membrane attachment to the cortical actin cytoskeleton is loosened by SigD/SopB, an inositol phosphatase that acts on PtdIns(4,5) P_2 . SopE and SopE2 enhance Cdc42 and Rac1 activity directly by acting as guanine-nucleotide-exchange factors. SipA and SipC alter cytoskeletal structure, SipC by nucleating actin and initiating polymerization and SipA by binding actin and modulating actin bundling. These cytoskeletal rearrangements are downregulated by the GAP (GTPase-activating protein) activity of SptP, which inactivates Cdc42 and Rac. SigD also is involved in sealing invaginating regions of the plasma membrane to form intracellular vacuoles. **b**, *Shigella* invasion. IpgD loosens membrane/cytoskeletal attachments in a mechanism similar to *Salmonella* SigD. VirA binds tubulin and promotes microtubule destabilization. This stimulates increased microtubule growth and causes Rac1 activation. Rho and Cdc42 are also activated during *Shigella* entry, through mechanisms that are still unclear. Activation of the small GTPases triggers filopodia and lamellipodia formation in the vicinity of the bacteria. IpaA binds vinculin and is involved in the transformation of filopodial extensions into membrane leaflets, and IpgD is probably involved in vesicle sealing. **c**, *Trypanosoma cruzi* invasion. Binding of an unidentified parasite product to host cells causes an increase in intracellular calcium. This stimulates increased microtubule growth and causes Rac1 activation. Activation of the small GTPases triggers filopodia and lamellipodia formation in the vicinity of the bacteria. IpaA binds vinculin and is involved in the transformation of filopodial extensions into membrane leaflets, and IpgD is probably involved in vesicle sealing. *Trypanosoma cruzi* uses this membrane to form a vacuole.

deplete PtdIns(4,5) P_2 (ref. 13) to mediate entry into non-phagocytic cells (Fig. 1b). However, compared to *Salmonella*, *Shigella* utilizes an additional means to affect the actin cytoskeleton. The *Shigella* effector VirA binds tubulin and promotes microtubule destabilization¹⁴. This stimulates increased microtubule growth, which has recently been shown to activate Rac1 in other systems¹⁵. Indeed, injection of VirA directly into cells induces membrane ruffling, which can be inhibited by dominant-negative Rac1 (ref. 14). The molecular mechanisms linking microtubule growth to Rac activation are still being elucidated and VirA may be a valuable probe for such studies.

Viruses enter cells through fusion of their membrane with the host-cell plasma membrane or via endocytosis¹⁶. But this is not a passive process, as binding of viral proteins to cellular receptors can initiate specific signalling cascades¹⁷. Several viruses bind to and initiate signalling through integrins to induce cellular uptake¹⁷, and although the role of actin in endocytosis is still controversial¹⁸, actin rearrangements are necessary for the entry of several viruses, including adenovirus and certain forms of vaccinia¹⁶. In addition, emerging evidence shows that many pathogens enter host cells through caveolae or lipid rafts, and viruses are no exception¹⁹. Studies of SV40 entry into cells have recently revealed a new microtubule-dependent transport pathway from plasma-membrane caveolae to the endoplasmic reticulum, the details of which are still under investigation²⁰. As such, the study of viral entry into host cells represents a fertile area for future discoveries.

The significantly larger protozoan parasites such as *Plasmodium falciparum* and *Toxoplasma gondii* do not rely on host-cell actin-dependent internalization machinery for host-cell invasion. Instead, they utilize an actinomyosin motor present in their own cytoskeleton to generate the motile force necessary to actively propel themselves inside cells^{21,22}. Host-cell actin polymerization is necessary, however, for the invasion process of some protozoan parasites such as *Cryptosporidium parvum*²³, but nothing is known about the mechanisms used by this parasite to initiate its cellular uptake. *Trypanosoma cruzi* invasion is actually enhanced by inhibitors of actin polymerization. This parasite uses a novel calcium-regulated microtubule-mediated pathway that directs recruitment and fusion of lysosomes to the plasma membrane of host cells. *T. cruzi* then co-opts this lysosomal-derived membrane to form a parasitophorous vacuole inside the host cell²⁴. Binding of *T. cruzi* to cells causes a transient increase in intracellular calcium that induces actin disruption and the mobilization of lysosomes towards the site of parasite attachment, and this is mediated by microtubules and associated kinesin motors (Fig. 1c). Studies in uninfected host cells have subsequently revealed that the calcium-induced fusion of lysosomes to the plasma membrane comprises a previously undiscovered pathway to recruit new membrane to the plasmalemma during wound repair²⁴.

Redirecting traffic

Once inside a host cell, an intracellular pathogen must use a strategy to avoid or withstand the maturation of its vacuole into a phagolysosome. Some pathogens have adapted to resist and thrive in the harsh phagolysosomal environment, others lyse their vacuole and escape to the cytoplasm, but many actively modify the vacuole to suit their needs²⁵. Vacuole remodelling by pathogens is an active process that involves both blocking of fusion with certain compartments, and promotion of fusion with others. Because the cytoskeleton is involved in membrane traffic events including phagosomal maturation, it follows that manipulation of the cytoskeleton is a potential means for pathogens to influence these processes. Indeed, the well-studied example of *Salmonella*, where some of the bacterial effectors and host factors involved have now been identified, reveals that both microtubules and actin are used by this pathogen to alter its vacuolar fate (Fig. 2a).

The *Salmonella*-containing vacuole (SCV) of epithelial cells and macrophages is a dynamic structure that undergoes a maturation process involving fusion with certain endosomal compartments, while avoiding fusion with others. *Salmonella* vacuole remodelling

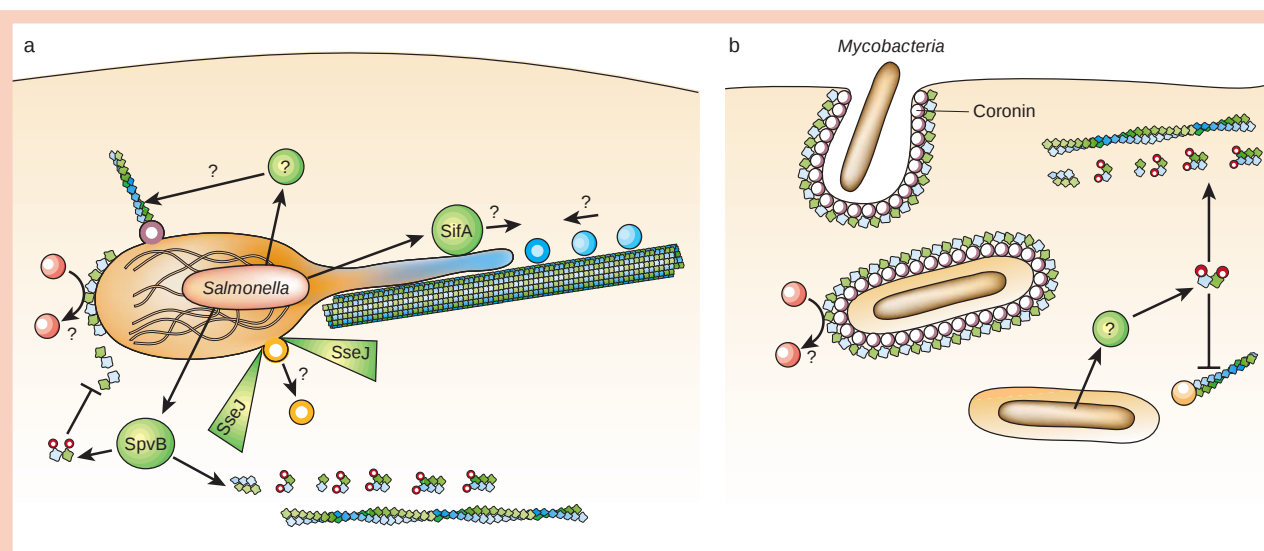


Figure 2 Vacuolar remodelling. **a**, *Salmonella*. Intracellular *Salmonella* reside and replicate within a vacuole known as the SCV (*Salmonella*-containing vacuole). Maintenance of this vacuole as a permissive environment for the bacteria involves altering normal vacuolar transport pathways. SifA is involved in the formation of tubular extensions to the SCV (Sifs) which, are involved in maintaining vacuolar integrity. Sifs form along microtubules and it is likely that *Salmonella* uses microtubules and their associated motor proteins to deliver membrane vesicles destined for fusion to the SCV, or to send out membranous SCV tentacles that fuse with vesicular compartments. An unknown SPI-2 effector mediates recruitment of actin to the SCV and is also involved in maintaining vacuolar integrity. This may involve recruitment and fusion of actin-containing or actin-propelled vesicles to the SCV. An actin coat may also protect the SCV from fusion with unfavourable compartments. SpvB modifies actin by ADP-ribosylation, and may be involved in a subsequent disassembly of actin at the SCV and other cellular sites, such as stress fibres. SseJ may be involved in membrane removal from the SCV, through budding or scission. **b**, *Mycobacteria*. Coronin is recruited to mycobacterial vacuoles during their formation. Retention of coronin on the vacuolar membrane may serve to protect the mycobacterial vacuole from undergoing unfavourable fusion events. *Mycobacteria* also disrupt the actin cytoskeleton, through an unknown mechanism. This disruption causes the vacuole to arrest during phagosomal maturation at a stage where it can fuse with early endosomes but not lysosomes. The mechanisms involved in this maturation arrest have not been described, but may involve downregulation of actin-mediated vesicle rocketing.

requires the concerted effort of several effector proteins of a second TTSS, encoded within SPI-2. *Salmonella* induce the formation of tubular membranous structures adjoining the SCV that are known as *Salmonella*-induced filaments or Sifs²⁶. Sif formation requires SifA²⁷, a bacterial effector protein translocated into host cells by the SPI-2 TTSS, and transfection of SifA into uninfected cells induces the formation of Sif-like structures²⁸. *sifA* mutants lose their vacuolar membrane and are released into the cytosol, leading to attenuation of *sifA* mutants in macrophages and mouse models^{29,30}. Sifs form along microtubules and the disruption of microtubules with nocodazole blocks Sif formation and inhibits *Salmonella* replication^{26,31}. Together, these results indicate that SifA is actively involved in the recruitment of membrane to the SCV and that this acquisition is microtubule dependent. It is likely that *Salmonella* uses microtubules and their associated motor proteins to deliver membrane vesicles destined for fusion to the SCV, or to send out membranous SCV tentacles that fuse with vesicular compartments.

In addition to its role in invasion, actin is also used in SCV remodelling by *Salmonella*. Approximately four hours after bacterial uptake, *Salmonella* induces the formation of an actin meshwork around the SCV³². This event requires the SPI-2 TTSS, although the translocated effectors involved have not yet been identified. Treatment of infected cells with actin-depolymerizing agents inhibits *Salmonella* replication in macrophages and results in loss of the SCV membrane and the release of bacteria into the cytoplasm³². SPI-2 directs actin assembly and Sif formation at the SCV, but only bacteria with an intact SPI-2 TTSS require actin and SifA to maintain vacuolar integrity^{29,32}. This suggests that a complex, sequential series of events is involved in SCV remodelling. At later times, *Salmonella* can mediate disruption of actin around the SCV and at other host-cell sites³¹, a phenotype that is attributed to *spvB*, which ADP-ribosylates actin³³. Another effector protein of the SPI-2 TTSS, *SseJ*, is proposed to be involved in budding or scission of the membrane from the SCV³⁰. Therefore, it appears that

remodelling and maintenance of the SCV membrane by *Salmonella* requires coordinated regulation of membrane acquisition and removal, and involves both microtubules and actin.

The role of actin in maintaining the SCV membrane remains to be elucidated, but recent papers have described a role for actin polymerization in movement and/or fusion of several endomembrane vesicle systems³⁴. This suggests that actin may be used for the recruitment and/or fusion of membranous compartments to the SCV. As *Salmonella* replicate, the amount of SCV membrane needs to increase to accommodate a growing population of bacteria, and this is probably accomplished by fusion of vesicles to the SCV. The nature of the compartments fusing with the SCV at later time points is a matter of controversy and remains to be determined. However, the recent observations that the SCV recruits actin³² and accumulates large amounts of cholesterol³⁶ at late time points may provide important clues to the mechanisms of SCV remodelling. Actin-mediated motility of endocytic and Golgi-derived vesicles has been described in fibroblast cells and is preferentially induced on membranes enriched in cholesterol/sphingolipid microdomains³⁵. Together, these observations suggest that actin-mediated vesicle rocketing may be involved in the recruitment and fusion of cholesterol-rich vesicles to the SCV. In contrast to a role in promoting vesicle fusion, in certain cases actin filaments may also be important in blocking fusion between particular compartments³⁷, raising the possibility that an actin meshwork may be involved in the SCV's avoidance of fusion with NADPH-oxidase-containing vesicles³⁸ or other compartments.

Vacuolar remodelling has been most extensively described for *Salmonella*, but as we learn more about the mechanisms used by other pathogens, it is likely that hijacking of the cytoskeleton for the manipulation of membrane transport will emerge as a common theme. Studies of *Mycobacteria* have also linked a pathogen-directed disruption of vacuolar transport with the actin cytoskeleton (Fig. 2b). Vacuoles containing virulent *Mycobacteria* do not fully

mature into phagolysosomes, and appear arrested during phagosomal maturation at a stage where they can fuse with early endosomes but not lysosomes. This maturation arrest occurs simultaneously with a bacterially induced disruption of the host-cell actin cytoskeleton, suggesting that these two events could be related³⁹. Indeed, disruption of actin filaments with inhibitors can induce a similar maturation arrest in latex-bead phagosomes⁴⁰. A second link between mycobacterial vacuole maturation and the actin cytoskeleton has also been described: the actin-binding protein coronin is recruited to and retained on mycobacterial vacuoles and has been implicated in their inability to fuse with other compartments⁴¹, although this is controversial⁴². Further characterization of these events should provide some insights into the relationship between the cytoskeleton and membrane transport.

Hitching a ride

As opposed to intracellular pathogens that live in membrane-bound vacuoles, several pathogens survive and replicate in the cell cytosol. These include pathogens that actively lyse their vacuole and those that enter the cytosol directly or through extrusion from an endosome. Once inside the cytosol, many of these pathogens harness and utilize the host cell's cytoskeletal machinery to move around.

Listeria, *Shigella* and vaccinia virus have provided cell biologists with valuable tools to study actin-based transport processes^{43,44}. All use actin-based motility to move within cells and/or spread between cells (Fig. 3). Enteropathogenic *Escherichia coli* (EPEC) represents a variation on this theme. It directs actin polymerization from an extracellular position to form an actin pedestal underneath extracellular adherent bacteria. Studies using cell extracts, reconstituted purified components and, more recently, cells from knockout mice genetically null for various cytoskeletal or signalling proteins, have facilitated detailed dissection of the roles of many components involved in actin polymerization. What is striking about the motility of these pathogens is that they have independently evolved mechanisms to harness the activity of the actin cytoskeleton at different points, yet their strategies all converge on the Arp2/3 complex.

The Arp2/3 complex is a seven-protein complex that, when activated, nucleates *de novo* actin polymerization⁴⁵. Arp2/3 is activated by the Wiskott–Aldrich syndrome protein (WASP) family, which consists of haematopoietic cell-specific WASP, ubiquitous neuronal WASP (N-WASP) and three Scar/WAVE family proteins. WASP proteins serve a scaffolding function to bring together actin monomers and Arp2/3 to form a nucleation core, the rate-limiting step in actin polymerization. N-WASP exists in an autoinhibited conformation, but is activated by the binding of a variety of proteins and/or lipids that relieve the autoinhibition or serve an accessory function to increase the rate of actin polymerization by Arp2/3 (ref. 45). The *Listeria* protein ActA binds the Arp2/3 complex directly, and has additional domains that bind actin monomers and vasodilator-stimulated phosphoprotein (VASP), thus mimicking the scaffolding function of N-WASP to activate Arp2/3 (refs 43,44). In contrast, IcsA/VirG of *Shigella* binds and activates N-WASP, which in turn recruits and activates the Arp2/3 complex^{43,44}. EPEC Tir, a TTSS-translocated protein of EPEC that is inserted into the host-cell plasma membrane, and the vaccinia protein A36R, are tyrosine phosphorylated in host cells and then bind the adaptor protein Nck, which recruits and activates N-WASP, which then recruits Arp2/3^{46,47}.

Thus, microbes have intersected the pathway from tyrosine phosphorylation to actin polymerization at virtually each step, providing a 'nested set' of reagents to study actin polymerization. Two other pathogens that show tantalizing promise for additional discovery are *Rickettsia*, which has actin-based motility that seems to be independent of Arp2/3 (ref. 48), and enterohaemorrhagic *E. coli*, which forms an actin pedestal without tyrosine phosphorylation or Nck^{43,44}. Future work on these pathogens should reveal more about alternate pathways that can induce actin polymerization.

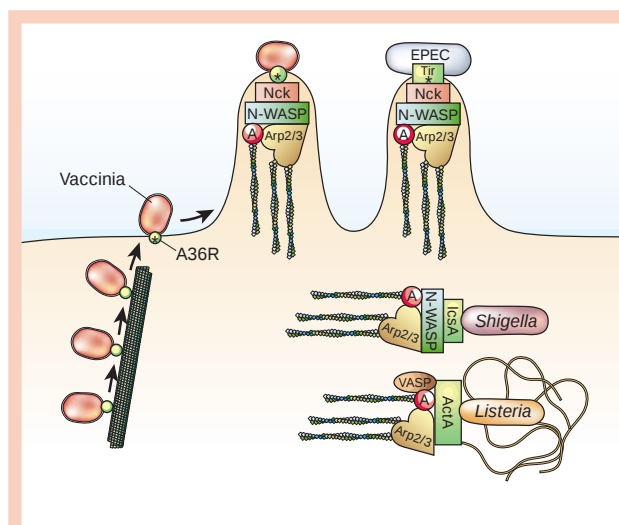


Figure 3 Intracellular transport. Vaccinia virus reaches the plasma membrane by transport on microtubules. At the plasma membrane, the viral protein A36R is tyrosine-phosphorylated, and this forms a binding site for the SH2 domain of the adaptor protein Nck. Nck recruitment to vaccinia leads to subsequent recruitment and activation of neuronal Wiskott–Aldrich syndrome protein (N-WASP), which brings together Arp2/3 and an actin monomer to initiate actin polymerization, directing the virus away from the cell. Enteropathogenic *Escherichia coli* (EPEC) inserts Tir into the plasma membrane, where it is tyrosine-phosphorylated and directs a recruitment cascade remarkably similar to vaccinia. This leads to the formation of an actin pedestal underneath adherent EPEC, the purpose of which is not known. *Shigella* IcsA binds and activates N-WASP, which facilitates binding of Arp2/3, while *Listeria* ActA mimics N-WASP, binding Arp2/3, actin monomers and vasodilator-stimulated phosphoprotein (VASP) directly. *Shigella* and *Listeria* use actin-based motility for transport through the cytoplasm and from cell to cell.

While it was previously thought that vaccinia use actin-based motility to propel themselves through the cytosol of infected cells, recent, more detailed observations have revealed that vaccinia actually utilize microtubules and associated kinesins for transport to the cell periphery and then switch to actin-based motility at the plasma membrane, where Src family kinases phosphorylate A36R to create a Nck-binding site⁴⁹. It is not entirely unexpected that vaccinia uses microtubules and motor proteins for directed transport to the cell periphery, as several other viruses also use microtubules for transport¹⁶. Herpesviruses, adenoviruses, and HIV all use minus-end-directed transport along microtubules to get from their point of entry at the cell periphery to their site of replication at the nucleus, and vaccinia most likely uses minus-end-directed transport to reach its perinuclear replication centre as well. After replication, plus-end-directed transport is used by several viruses for directed movement to the cell periphery for release. Some viruses seem to engage both plus- and minus-end-directed motors simultaneously, regulating their activity to favour motion in the desired direction. Future work aimed at understanding the mechanisms viruses use to engage and regulate molecular motors should shed light on motor function.

The path of least resistance

Many pathogens, including bacteria, viruses and protozoan parasites disrupt tight junctions during infection⁵⁰. Tight junctions seal the space between adjacent cells, limiting diffusion of solutes through the intercellular space and creating a boundary between the apical and basolateral sides of cellular barriers such as epithelia⁵¹. Tight junctions consist of integral membrane proteins (such as occludin, claudins and junctional adhesion molecule) and cytoplasmic PDZ-domain-containing proteins (zonula occludens (ZO)-1, -2, -3, MAGUK family and PAR family), the latter of which act as adaptors at the cytoplasmic

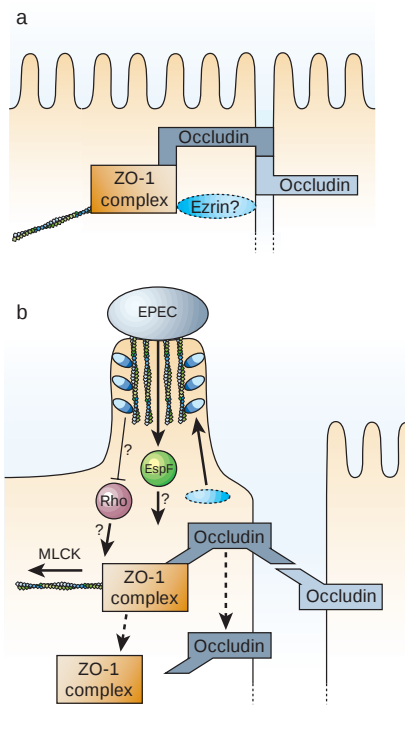


Figure 4 Disruption of tight junctions. **a**, Tight junctions consist of integral membrane proteins and cytoplasmic PDZ-domain-containing proteins that act as adaptors at the cytoplasmic surface, binding each other, as well as occludin, claudins, actin and several cytoplasmic proteins. Extracellular domains of tight junction membrane proteins from one cell bind those on adjacent cells, forming the basis of the intercellular seal. Zonula occludens (ZO)-1 and -2 link tight junctions to the actin cytoskeleton by binding the tight junction transmembrane proteins and actin. Ezrin may be important directly or indirectly in tight junction integrity (see text). **b**, Enteropathogenic *Escherichia coli* (EPEC)-induced disruption of tight junctions requires the type III secretion system (TTSS-translocated effector EspF. The mechanism of action of EspF remains to be determined, but disruption of tight junctions by EPEC is associated with a variety of effects: myosin light chain (MLC) and ezrin phosphorylation, occludin dephosphorylation, and dissociation of occludin and ZO-1 from tight junctions. MLC is phosphorylated by myosin light chain kinase (MLCK), which causes contraction of the perijunctional actomyosin ring. This opens tight junctions, presumably by increasing the tension on the tight junction, resulting in a decrease of the transepithelial resistance. Occludin is dephosphorylated and occludin and ZO-1 dissociate from tight junctions. EPEC's effects on tight junctions may be mediated by ezrin, either by sequestration of ezrin away from tight junctions by its recruitment to the pedestal, or ezrin effects on signalling through Rho-family proteins (see text).

surface of tight junctions, binding each other, as well as occludin, claudins, actin and several cytoplasmic proteins. ZO-1 and ZO-2 link tight junctions to the actin cytoskeleton by binding the tight junction transmembrane proteins and actin with their N- and C termini, respectively. This complex links tight junctions to a perijunctional actomyosin ring, which supports and regulates tight junction permeability. The strategies used by pathogens for altering tight junction permeability are as numerous as the pathogens themselves⁵⁰. Whereas some pathogens bind and modify junction components directly, others exert their effects through the actin cytoskeleton, which controls the integrity of tight junctions.

Rho family proteins are implicated in the assembly and maintenance of tight junctions, and many bacteria produce toxins that modify these small GTP-binding proteins, leading to disruption of tight junctions⁵². One example is the diarrhoeagenic pathogen

Clostridium difficile, which produces two toxins (A and B) that inactivate RhoA by glucosylation. This causes F-actin restructuring, dissociation of actin from ZO-1, dissociation of occludin, ZO-1 and ZO-2 from the tight junction, and a decrease in transepithelial resistance of epithelial monolayers⁵⁰.

Enteropathogenic *E. coli* also targets tight junctions through manipulation of the actin cytoskeleton. Although much work has been done to identify the host and bacterial players involved in this phenomenon, the exact sequence of events and their relation to each other is not yet fully understood. Decreases in transepithelial resistance during EPEC infection can be correlated with myosin light chain (MLC) phosphorylation⁵³, ezrin phosphorylation⁵⁴, occludin dephosphorylation⁵⁵ and occludin⁵⁵ and ZO-1 (ref. 56) dissociation from tight junctions (Fig. 4). MLC phosphorylation causes contraction of the perijunctional actomyosin ring, which opens tight junctions, presumably by increasing the tension on the tight junction, resulting in a decrease of the transepithelial resistance. Inhibitors of MLC kinase partially prevent the decrease in transepithelial resistance seen during EPEC infection⁵³.

Ezrin is a member of the closely related ezrin–radixin–moesin (ERM) family of proteins that mediate membrane–cytoskeletal linkages. Phosphorylation and increased cytoskeletal association of ezrin are seen during EPEC infection, and transfection of dominant-negative ezrin partially blocks the effects of EPEC on tight junctions⁵⁴. The relationship between ezrin and tight junctions in normal or EPEC-infected cells is currently not well defined. Ezrin is one of the main components of EPEC actin pedestals⁵⁷, which form at the apical surface of infected cells underneath adherent EPEC. An attractive hypothesis is that some proportion of ezrin localizes to tight junctions, and is sequestered away from this site by its recruitment to the EPEC pedestal, although additional factors are required (see below). Alternatively, EPEC's use of ezrin to disrupt tight junctions may involve signalling through Rho. Ezrin has been implicated in signalling both upstream and downstream of Rho, and recent studies in *Drosophila* demonstrate that ERM proteins control epithelial polarity and integrity through antagonistic effects on Rho signalling⁵⁸. However, the mechanisms of cell–cell adhesion in *Drosophila* are distinct from those in mammals, as *Drosophila* do not have tight junctions. Clarification of ezrin's role in the dynamics of mammalian tight junctions awaits further experimentation, and EPEC may be a key tool in unravelling this relationship.

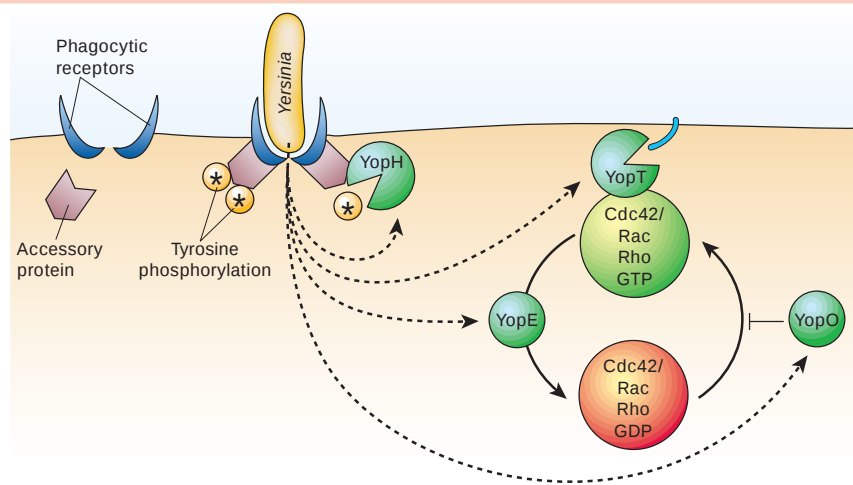
The bacterial protein implicated in the effects of EPEC on tight junctions is a TTSS effector protein called EspF⁵⁹. EspF is required in a dose-dependent manner for the decrease in transepithelial resistance and occludin redistribution seen during EPEC infection⁵⁹, and ezrin activation is attenuated in an EPEC strain lacking EspF⁵⁴. Little is known about the function of EspF, but it contains proline-rich domains that resemble Src-homology 3 (SH3)-binding or tryptophan-tryptophan (WW)-binding domains, suggesting that it may interact with one or more host-cell proteins via these domains. Several components of tight junctions have SH3 domains, but no binding partners have yet been described for EspF. Identification of EspF binding partners and precise subcellular localization of EspF inside host cells should clarify the manner in which this bacterial effector affects tight junctions.

Polymorphonuclear disarmament

In contrast to activation of cytoskeletal remodelling for entry into host cells, some pathogens paralyse the cytoskeleton to avoid uptake by phagocytes. Phagocytic cells such as macrophages and polymorphonuclear cells (neutrophils) possess several pathways of phagocytosis². Opsonin-dependent pathways are mediated by FcγR and complement receptors, and opsonin-independent uptake can be mediated by a variety of receptors, including scavenger receptors, mannose receptors and integrins. Regardless of the specific pathways or particles involved, all phagocytic processes are driven by complex, controlled rearrangements of the actin cytoskeleton²,

Figure 5 *Yersinia* avoidance of phagocytosis.

Binding of *Yersinia* to host-cell receptors triggers phagocytic pathways that would normally result in bacterial uptake. The rapid translocation of several effectors by *Yersinia* disarms these pathways, facilitating bacterial avoidance of phagocytosis. *Yersinia* outer protein H (YopH) dephosphorylates a number of tyrosine-phosphorylated signalling proteins including Fyb, SKAP-HOM and p130^{cas}, thereby disrupting their abilities to mediate further downstream signalling events in the cytoskeletal pathway. YopE disrupts actin filaments by acting as a GTPase-activating protein for the GTPases Rac1, Rho and Cdc42, while YopT proteolytically cleaves this family of proteins, resulting in their release from the membrane. YopO blocks the activation of Rho through a mechanism that is not fully understood.



and disruption of the cytoskeleton is an effective means by which pathogens can block their own uptake.

Yersinia can trigger several key phagocytic pathways: in the absence of opsonization, *Yersinia* adherence is mediated by the adhesins YadA and invasin, which trigger internalization via β 1-integrins in phagocytic and non-phagocytic cells⁶⁰. Opsonization of *Yersinia* facilitates cell binding via Fc γ R and complement receptors in professional phagocytes⁶¹. Regardless of the adherence mechanism, *Yersinia* counteracts its own receptor binding and activation by injecting effector proteins that disrupt the actin cytoskeleton and defuse the triggered phagocytic pathways (Fig. 5). Remarkably, of the six known effectors of the *Yersinia* TTSS, four target the actin cytoskeleton and contribute to blocking phagocytosis. *Yersinia* outer proteins (Yops) E, H, T and O all have so-called 'anti-phagocytic' effects⁶¹.

The first characterized anti-phagocytic effector of *Yersinia* is YopH, a protein tyrosine phosphatase⁶². *Yersinia* interaction with host cells leads to rapid phosphorylation of several host proteins that are then rapidly dephosphorylated by YopH. Several substrates have been identified for YopH: Crk-associated substrate (Cas), focal adhesion kinase (FAK), paxillin, Fyn-T-binding protein/SLP-76-associated protein/ADAP (Fyb/SLAP/ADAP), the scaffolding protein SKAP-HOM⁶³ and Pyk2 (ref. 64). Which of these substrates is involved in blocking phagocytosis has not been analysed systematically, but, conceivably, a role in actin rearrangements and phagocytic uptake could be proposed for each of them. Blocking either Cas or FAK activity decreases uptake of *Yersinia pseudotuberculosis* in fibroblasts⁶⁴, and FAK may be involved in Fc γ R-mediated phagocytosis, although this remains controversial². Paxillin is a FAK substrate that localizes to Fc γ R- and complement-mediated phagosomes and is strongly phosphorylated during Fc γ R-mediated uptake^{65,66}. Fyb/SLAP/ADAP and SKAP-HOM bind each other in a multi-protein complex that is implicated in integrin-mediated adhesion during T-cell signalling⁶⁷. A similar complex containing Fyb/SLAP/ADAP has recently been described in Fc γ R-mediated phagocytosis⁶⁸. The relevance of multiple YopH substrates has not been investigated thoroughly, but the observations that *Yersinia* uptake can occur via multiple phagocytic pathways, and that YopH can act in trans to inhibit phagocytosis through Fc γ R receptors⁶⁹, raise the possibility that multiple substrates of YopH facilitate the blockage of several pathways of phagocytosis simultaneously.

Not all phagocytic pathways require tyrosine phosphorylation. However, other anti-phagocytic Yops target the cytoskeleton through their effects on Rho family GTPases. Rho GTPases have key roles in phagocytosis, but the requirement for different family members varies with the mode of uptake². Complement receptor-mediated phagocytosis is Rho-dependent, whereas Fc γ R-mediated phagocytosis requires both Cdc42 and Rac. Rac1 has been implicated in

integrin-mediated uptake of *Yersinia* into HeLa cells, although other Rho family members may also be involved⁷⁰. To disarm these various pathways, *Yersinia* encodes three effectors that inactivate Rho GTPases in distinct ways. YopE is a GAP that has a broad specificity for Rho, Rac and Cdc42 *in vitro*⁷¹, although there is evidence that *in vivo* it may specifically inactivate either Rac or RhoA in certain contexts^{71,72}. Using a novel strategy for inactivating Rho-family proteins, YopT recognizes isoprenylated Rho, Rac and Cdc42 and cleaves them near the C terminus, resulting in their release from the membrane⁷³. YopO (called YpkA in *Yersinia pseudotuberculosis*) is a serine/threonine kinase that binds and is activated by actin, and it also interacts with Rho and Rac. YopO blocks the activation of RhoA in *Yersinia*-infected cells and has sequence similarity with eukaryotic Rho-binding kinases, but the mechanism by which YopO mediates its anti-phagocytic effect is not fully understood⁶⁰.

The presence of four anti-phagocytic Yops establishes *Yersinia* as an anti-phagocytic specialist, although several other pathogens also inhibit their own phagocytosis⁶³. At least two other pathogens act through the cytoskeleton to block their own uptake. *Pseudomonas aeruginosa* resists internalization through manipulation of Rho-family GTPases through mechanisms similar to *Yersinia*, while EPEC inhibits the phosphatidylinositol-3-OH kinase signalling pathways that are essential for actin polymerization and resultant phagocytic uptake⁶³.

Future directions

The collection of known microbial effectors provides an extensive set of reagents for cell biologists and those wishing to understand the physiological causes of infectious disease. The versatility of the cytoskeleton makes it a particularly attractive target for microbes, which use it as a multipurpose target to achieve many ends. As genetics, genomics and proteomics lead to the discovery of more microbial effectors, the number of reagents will continue to expand. Additionally, as our knowledge of cell biology increases, so will our ability to understand the activities of these effectors inside host cells. It is reasonable to speculate that for every host-cell cytoskeletal pathway, there is a microbial effector protein that exploits it, and the combined investigation of these pathways and their effectors should continue to foster insights that are mutually beneficial to both cell biologists and those studying infectious processes. □

doi:10.1038/nature01603

- Sandvig, K. & Van Deurs, B. Membrane traffic exploited by protein toxins. *Annu. Rev. Cell Dev. Biol.* **18**, 1–24 (2002).
- May, R. C. & Machesky, L. M. Phagocytosis and the actin cytoskeleton. *J. Cell Sci.* **114**, 1061–1077 (2001).
- Amyere, M. et al. Origin, originality, functions, subversions and molecular signalling of macropinocytosis. *Int. J. Med. Microbiol.* **291**, 487–494 (2002).

4. Hueck, C. J. Type III protein secretion systems in bacterial pathogens of animals and plants. *Microbiol. Mol. Biol. Rev.* **62**, 379–433 (1998).
5. Zhou, D. & Galán, J. *Salmonella* entry into host cells: the work in concert of type III secreted effector proteins. *Microbes Infect.* **3**, 1293–1298 (2001).
6. Buchwald, G. *et al.* Structural basis for the reversible activation of a Rho protein by the bacterial toxin SopE. *EMBO J.* **21**, 3286–3295 (2002).
7. Raucher, D. *et al.* Phosphatidylinositol 4,5-bisphosphate functions as a second messenger that regulates cytoskeleton-plasma membrane adhesion. *Cell* **100**, 221–228 (2000).
8. Terebiznik, M. R. *et al.* Elimination of host cell PtdIns(4,5)P₂ by bacterial SigD promotes membrane fission during invasion by *Salmonella*. *Nature Cell Biol.* **4**, 766–773 (2002).
9. Fu, Y. & Galán, J. E. A *Salmonella* protein antagonizes Rac-1 and Cdc42 to mediate host-cell recovery after bacterial invasion. *Nature* **401**, 293–297 (1999).
10. Murl, S., Watson, R. O. & Galán, J. E. Role of tyrosine kinases and the tyrosine phosphatase SptP in the interaction of *Salmonella* with host cells. *Cell. Microbiol.* **3**, 795–810 (2001).
11. Carlson, S. A., Omary, M. B. & Jones, B. D. Identification of cytokeratins as accessory mediators of *Salmonella* entry into eukaryotic cells. *Life Sci.* **70**, 1415–1426 (2002).
12. Tran Van Nhieu, G., Bourdet-Sicard, R., Duménil, G., Blocker, A. & Sansonetti, P. J. Bacterial signals and cell responses during *Shigella* entry into epithelial cells. *Cell. Microbiol.* **2**, 187–193 (2000).
13. Niebuhr, K. *et al.* Conversion of PtdIns(4,5)P₂ into PtdIns(5)P by the *S. flexneri* effector IgpD reorganizes host cell morphology. *EMBO J.* **21**, 5069–5078 (2002).
14. Yoshida, S. *et al.* *Shigella* deliver an effector protein to trigger host microtubule destabilization, which promotes Rac1 activity and efficient bacterial internalization. *EMBO J.* **21**, 2923–2935 (2002).
15. Waterman-Storer, C. M., Worthylake, R. A., Liu, B. P., Burridge, K. & Salmon, E. D. Microtubule growth activates Rac1 to promote lamellipodial protrusion in fibroblasts. *Nature Cell Biol.* **1**, 45–50 (1999).
16. Smith, G. A. & Enquist, L. W. BREAK INS AND BREAK OUTS: viral interactions with the cytoskeleton of mammalian cells. *Annu. Rev. Cell Dev. Biol.* **18**, 135–161 (2002).
17. Nemerow, G. R. & Cheres, D. A. Herpesvirus hijacks an integrin. *Nature Cell Biol.* **4**, E69–E71 (2002).
18. Schmid, S. L. & Sorkin, A. D. Days and nights discussing membrane dynamics in endocytosis: meeting report from the Euresco/EMBL Membrane Dynamics in Endocytosis, 6–11 October in Tomar, Portugal. *Traffic* **3**, 77–85 (2002).
19. van der Goot, F. G. & Harder, T. Raft membrane domains: from a liquid-ordered membrane phase to a site of pathogen attack. *Semin. Immunol.* **13**, 89–97 (2001).
20. Pelkmans, L., Kartenbeck, J. & Helenius, A. Caveolar endocytosis of simian virus 40 reveals a new two-step vesicular-transport pathway to the ER. *Nature Cell Biol.* **3**, 473–483 (2001).
21. Sibley, L. D. & Andrews, N. W. Cell invasion by un-palatable parasites. *Traffic* **1**, 100–106 (2000).
22. Cowman, A. F. & Crabb, B. S. The *Plasmodium falciparum* genome—a blueprint for erythrocyte invasion. *Science* **298**, 126–128 (2002).
23. Elliott, D. A. *et al.* *Cryptosporidium parvum* infection requires host cell actin polymerization. *Infect. Immun.* **69**, 5940–5942 (2001).
24. Tan, H. & Andrews, N. W. Don't bother to knock—the cell invasion strategy of *Trypanosoma cruzi*. *Trends Parasitol.* **18**, 427–428 (2002).
25. Méresse, S. *et al.* Controlling the maturation of pathogen-containing vacuoles: a matter of life and death. *Nature Cell Biol.* **1**, E183–E188 (1999).
26. Garcia-del Portillo, F., Zwick, M. B., Leung, K. Y. & Finlay, B. B. *Salmonella* induces the formation of filamentous structures containing lysosomal membrane glycoproteins in epithelial cells. *Proc. Natl Acad. Sci. USA* **90**, 10544–10548 (1993).
27. Stein, M. A., Leung, K. Y., Zwick, M., Garcia-del Portillo, F. & Finlay, B. B. Identification of a *Salmonella* virulence gene required for formation of filamentous structures containing lysosomal membrane glycoproteins within epithelial cells. *Mol. Microbiol.* **20**, 151–164 (1996).
28. Brummell, J. H., Rosenberger, C. M., Gotto, G. T., Marcus, S. L. & Finlay, B. B. SifA permits survival and replication of *Salmonella typhimurium* in murine macrophages. *Cell. Microbiol.* **3**, 75–84 (2001).
29. Beuzón, C. R. *et al.* *Salmonella* maintains the integrity of its intracellular vacuole through the action of SifA. *EMBO J.* **19**, 3235–3249 (2000).
30. Ruiz-Albert, J. *et al.* Complementary activities of SseI and SifA regulate dynamics of the *Salmonella typhimurium* vacuolar membrane. *Mol. Microbiol.* **44**, 645–661 (2002).
31. Brummell, J. H., Goosney, D. L. & Finlay, B. B. SifA, a type III secreted effector of *Salmonella typhimurium*, directs *Salmonella*-induced filament (Sif) formation along microtubules. *Traffic* **3**, 407–415 (2002).
32. Méresse, S. *et al.* Remodelling of the actin cytoskeleton is essential for replication of intravacuolar *Salmonella*. *Cell. Microbiol.* **3**, 567–577 (2001).
33. Lesnick, M. L., Reiner, N. E., Fierer, J. & Guiney, D. G. The *Salmonella* *spvB* virulence gene encodes an enzyme that ADP-ribosylates actin and destabilizes the cytoskeleton of eukaryotic cells. *Mol. Microbiol.* **39**, 1464–1470 (2001).
34. Taunton, J. Actin filament nucleation by endosomes, lysosomes and secretory vesicles. *Curr. Opin. Cell Biol.* **13**, 85–91 (2001).
35. Rozelle, A. L. *et al.* Phosphatidylinositol 4,5-bisphosphate induces actin-based movement of raft-enriched vesicles through WASP-Arp2/3. *Curr. Biol.* **10**, 311–320 (2000).
36. Catron, D. M. *et al.* The *Salmonella*-containing vacuole is a major site of intracellular cholesterol accumulation and recruits the GPI-anchored protein CD55. *Cell. Microbiol.* **4**, 315–328 (2002).
37. Muallem, S., Kwiatkowska, K., Xu, X. & Yin, H. L. Actin filament disassembly is a sufficient final trigger for exocytosis in nonexcitable cells. *J. Cell Biol.* **128**, 589–598 (1995).
38. Vazquez-Torres, A. *et al.* *Salmonella* pathogenicity island 2-dependent evasion of the phagocyte NADPH oxidase. *Science* **287**, 1655–1658 (2000).
39. Guerin, I. & de Chastellier, C. Pathogenic mycobacteria disrupt the macrophage actin filament network. *Infect. Immun.* **68**, 2655–2662 (2000).
40. Guerin, I. & de Chastellier, C. Disruption of the actin filament network affects delivery of endocytic contents marker to phagosomes with early endosome characteristics: the case of phagosomes with pathogenic mycobacteria. *Eur. J. Cell Biol.* **79**, 735–749 (2000).
41. Ferrari, G., Langen, H., Naito, M. & Pieters, J. A coat protein on phagosomes involved in the intracellular survival of mycobacteria. *Cell* **97**, 435–447 (1999).
42. Russell, D. G., Mwandumba, H. C. & Rhoades, E. E. Mycobacterium and the coat of many lipids. *J. Cell Biol.* **158**, 421–426 (2002).
43. Frischknecht, F. & Way, M. Surfing pathogens and the lessons learned for actin polymerization. *Trends Cell Biol.* **11**, 30–38 (2001).
44. Goldberg, M. B. Actin-based motility of intracellular microbial pathogens. *Microbiol. Mol. Biol. Rev.* **65**, 595–626 (2001).
45. Caron, E. Regulation of Wiskott-Aldrich syndrome protein and related molecules. *Curr. Opin. Cell Biol.* **14**, 82–87 (2002).
46. Frischknecht, F. *et al.* Actin-based motility of vaccinia virus mimics receptor tyrosine kinase signalling. *Nature* **401**, 926–929 (1999).
47. Gruenheid, S. *et al.* Enteropathogenic *E. coli* Tir binds Nck to initiate actin pedestal formation in host cells. *Nature Cell Biol.* **3**, 856–859 (2001).
48. Gouin, E. *et al.* A comparative study of the actin-based motilities of the pathogenic bacteria *Listeria monocytogenes*, *Shigella flexneri* and *Rickettsia conorii*. *J. Cell Sci.* **112**, 1697–1708 (1999).
49. Moss, B. & Ward, B. M. High-speed mass transit for poxviruses on microtubules. *Nature Cell Biol.* **3**, E245–E246 (2001).
50. Sears, C. L. Molecular physiology and pathophysiology of tight junctions V. Assault of the tight junction by enteric pathogens. *Am. J. Physiol. Gastrointest. Liver Physiol.* **279**, G1129–G1134 (2000).
51. Tsukita, S., Furuse, M. & Itoh, M. Multifunctional strands in tight junctions. *Nature Rev. Mol. Cell Biol.* **2**, 285–293 (2001).
52. Steele-Mortimer, O., Knodler, L. A. & Finlay, B. B. Poisons, ruffles and rockets: bacterial pathogens and the host cell cytoskeleton. *Traffic* **1**, 107–118 (2000).
53. Yuh, R., Koutsouris, A., Savkovic, S. D. & Hecht, G. Enteropathogenic *Escherichia coli*-induced myosin light chain phosphorylation alters intestinal epithelial permeability. *Gastroenterology* **113**, 1873–1882 (1997).
54. Simonovic, I., Arpin, M., Koutsouris, A., Falk-Krzesinski, H. J. & Hecht, G. Enteropathogenic *Escherichia coli* activates ezrin, which participates in disruption of tight junction barrier function. *Infect. Immun.* **69**, 5679–5688 (2001).
55. Simonovic, I., Rosenberg, J., Koutsouris, A. & Hecht, G. Enteropathogenic *Escherichia coli* dephosphorylates and dissociates occludin from intestinal epithelial tight junctions. *Cell. Microbiol.* **2**, 305–315 (2000).
56. Philpott, D. J., McKay, D. M., Sherman, P. M. & Perdue, M. H. Infection of T84 cells with enteropathogenic *Escherichia coli* alters barrier and transport functions. *Am. J. Physiol.* **270**, G634–G645 (1996).
57. Goosney, D. L., DeVinney, R. & Finlay, B. B. Recruitment of cytoskeletal and signaling proteins to enteropathogenic and enterohemorrhagic *Escherichia coli* pedestals. *Infect. Immun.* **69**, 3315–3322 (2001).
58. Speck, O., Hughes, S. C., Noren, N. K., Kulikauskas, R. M. & Fehon, R. G. Moesin functions antagonistically to the Rho pathway to maintain epithelial integrity. *Nature* **421**, 83–87 (2003).
59. McNamara, B. P. *et al.* Translocated EspF protein from enteropathogenic *Escherichia coli* disrupts host intestinal barrier function. *J. Clin. Invest.* **107**, 621–629 (2001).
60. Cornelis, G. R. *Yersinia* type III secretion: send in the effectors. *J. Cell Biol.* **158**, 401–408 (2002).
61. Grosdent, N., Maridonneau-Parini, I., Sory, M. P. & Cornelis, G. R. Role of Yops and adhesins in resistance of *Yersinia enterocolitica* to phagocytosis. *Infect. Immun.* **70**, 4165–4176 (2002).
62. Zhang, Z. Y. *et al.* Expression, purification, and physicochemical characterization of a recombinant *Yersinia* protein tyrosine phosphatase. *J. Biol. Chem.* **267**, 23759–23766 (1992).
63. Celli, J. & Finlay, B. B. Bacterial avoidance of phagocytosis. *Trends Microbiol.* **10**, 232–237 (2002).
64. Bruce-Staskal, P. J., Weidow, C. L., Gibson, J. J. & Bouton, A. H. Cas, Fak and Pyk2 function in diverse signaling cascades to promote *Yersinia* uptake. *J. Cell Sci.* **115**, 2689–2700 (2002).
65. Greenberg, S., Chang, P. & Silverstein, S. C. Tyrosine phosphorylation of the γ subunit of Fc γ receptors, p72^{ph}, and paxillin during Fc receptor-mediated phagocytosis in macrophages. *J. Biol. Chem.* **269**, 3897–3902 (1994).
66. Allen, L. A. & Aderem, A. Molecular definition of distinct cytoskeletal structures involved in complement- and Fc receptor-mediated phagocytosis in macrophages. *J. Exp. Med.* **184**, 627–637 (1996).
67. Griffiths, E. K. & Penninger, J. M. Communication between the TCR and integrins: role of the molecular adapter ADAP/Fyb/Slap. *Curr. Opin. Immunol.* **14**, 317–322 (2002).
68. Coppolino, M. G. *et al.* Evidence for a molecular complex consisting of Fyb/SLAP, SLP-76, Nck, VASP and WASP that links the actin cytoskeleton to Fc γ receptor signalling during phagocytosis. *J. Cell Sci.* **114**, 4307–4318 (2001).
69. Fallman, M. *et al.* *Yersinia pseudotuberculosis* inhibits Fc receptor-mediated phagocytosis in J774 cells. *Infect. Immun.* **63**, 3117–3124 (1995).
70. McGee, K., Zetti, M., Way, M. & Fallman, M. A role for N-WASP in invasive-promoted internalisation. *FEBS Lett.* **509**, 59–65 (2001).
71. Black, D. S. & Bliska, J. B. The RhoGAP activity of the *Yersinia pseudotuberculosis* cytotoxin YopE is required for antiphagocytic function and virulence. *Mol. Microbiol.* **37**, 515–527 (2000).
72. Andor, A. *et al.* YopE of *Yersinia*, a GAP for Rho GTPases, selectively modulates Rac-dependent actin structures in endothelial cells. *Cell. Microbiol.* **3**, 301–310 (2001).
73. Shao, F., Merritt, P. M., Bao, Z., Innes, R. W. & Dixon, J. E. A *Yersinia* effector and a *Pseudomonas* avirulence protein define a family of cysteine proteases functioning in bacterial pathogenesis. *Cell* **109**, 575–588 (2002).

Acknowledgements We apologize to the many scientists whose work we could not discuss or cite directly owing to space limitations, and we thank D. Goosney, J. Brummell, G. Hecht, M. Grigg and members of the Finlay lab for comments on the manuscript. S.G. is supported by a postdoctoral fellowship from the Canadian Institutes of Health Research (CIHR) and B.B.F. is a CIHR Distinguished Investigator, a Howard Hughes Medical Institute International Research Scholar, and the University of British Columbia Peter Wall Distinguished Professor.



CYTOKINETICS / GlaxoSmithKline
COLLABORATION

Leading the way to the next generation of anti-mitotics

Historical experience has demonstrated that important progress is often made in clinical medicine with the introduction of pharmaceuticals that act by novel mechanisms of action. Such is the case in the area of anti-cancer therapy. Existing anti-mitotics (taxanes and vinca alkaloids) are all directed at tubulin, the intracellular protein that comprises the mitotic spindle, and are perhaps the most clinically and commercially successful anti-cancer agents. Since their introduction 20 years ago, these agents have dramatically advanced cancer patient care and have served as a cornerstone of modern chemotherapy.

However, use of these agents can be constrained by dose-limiting toxicities related to the broad role tubulin plays in important cellular processes unrelated to mitosis. In contrast, mitotic kinesins represent a family of newly identified enzymes, each of which appears to perform discrete and non-redundant roles in mitotic spindle formation and function during cell division. Unlike tubulin, mitotic kinesins are expressed only in proliferating cells and appear to play no role outside of mitosis. Inhibition of mitotic kinesins disrupts the cell cycle, thereby inducing apoptosis or cell death. Inhibitors of mitotic kinesins may therefore



represent the next generation of anti-mitotics; they target a new set of molecular enzymes specifically involved in the mitotic process, yet within a well-validated area of pharmaceutical development. Because mitotic kinesin inhibitors differ from existing anti-mitotic drugs in their

molecular targets and mechanism of action, they are being investigated for their potential therapeutic profile. In April 2002, at the annual meeting of the American Association for Cancer Research, Cytokinetics and GlaxoSmithKline jointly unveiled the results of intensive research efforts in the field of mitotic kinesins, which covered the breadth of anti-

tumor activity observed for kinesin spindle protein (KSP) inhibitors in multiple preclinical studies of cancer and demonstrated the absence of neuropathy in animal models. Based on these preclinical findings, Cytokinetics and GlaxoSmithKline initiated a broad-based clinical development program with the first of several anticipated pharmaceutical candidates that are direct inhibitors of mitotic kinesin targets.

Kenneth W. Wood, Ph.D.

David J. Morgans, Jr., Ph.D.

James H. Sabry, M.D., Ph.D.

Jeffrey R. Jackson, Ph.D.

John D. Elliott, Ph.D.

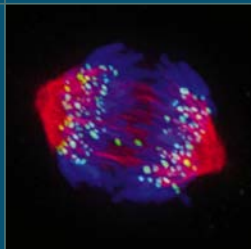
Pearl S. Huang, Ph.D.



CYTOKINETICS



GlaxoSmithKline



About Cytokinetics

Cytokinetics is dedicated to the discovery, development and commercialization of novel therapeutics within the emerging field of cytoskeletal pharmacology. The cytoskeleton is a complex, dynamic framework that impacts all aspects of cell function including division, motility, transport, muscle contractility and cellular organization. Cytokinetics' R&D efforts aim to address pharmaceutical needs in cancer, cardiovascular and infectious diseases and feature proprietary Cytometrix™ cellular phenotyping technologies designed to industrialize cell biology for increased speed and productivity in drug discovery and development. Cytokinetics and GlaxoSmithKline have begun Phase I studies with the first novel anti-cancer drug candidate under their collaboration. The company is building a promising pipeline of next-generation drug candidates arising from other internal pharmaceutical research programs and is planning the initiation of clinical studies for other potential first-in-class compounds.

Cytokinetics' mission is to translate the power of cell biology into novel pharmaceuticals and technologies with the objective to enhance patient care.



CYTOKINETICS

For more information, visit Cytokinetics on the World Wide Web at www.cytokinetics.com

Contacts

Publisher: Ben Crowe

Editor: Paul Smaglik

Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building

4 Crinan Street

London N1 9XW, UK

Tel +44 (0) 20 7843 4961

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Senior European Sales Manager:

Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)

Andy Douglas (4975)

Frank Phelan (4944)

Netherlands/ Italy/ Iberia:

Evelina Rubio Håkansson (4973)

Scandinavia: Silje Opstrup (4994)

France/ Belgium:

Amelie Pequignot (4974)

Production Manager:

 Billie Franklin

To send materials use London

address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Hind Berrada (4935)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Office

Germany/ Austria/ Switzerland:

Patrick Phelan, Odo Wulffen

Tel + 49 89 54 90 57 11/-2

Fax + 49 89 54 90 57 20

e-mail: p.phelan@nature.com

o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,

10th Floor, New York, NY 10010-1707

Tel +1 800 989 7718

Fax +1 800 989 7103

e-mail: naturejobs@natureny.com

US Sales Manager:

 Peter Bless

US Advertising Coordinator:

Linda Adam

Japan Head Office, Tokyo

MG Ichigaya Building (6F),

19-1 Harakikatamachi,

Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

Asia-Pacific Sales Director:

Hideki Watanabe

e-mail: h.watanabe@naturejpn.com

naturejobs

Salary isn't everything

Richard Sykes, university administrator and devil's advocate, wonders why a student graduating from his institution would ever want to go on to do a PhD. Other options are much more palatable, says the rector of Imperial College, London, tongue planted at least half in his cheek. "You can go to the City and earn about £50,000 (US\$78,000) a year," he notes. Whereas doing a PhD in Britain means enduring years of scant financial support and increased competition — even for lecturer positions. "The prospect is not very encouraging," Sykes says.

Sykes has repeatedly called for higher salaries at UK universities, which would help to make scientific careers more attractive. But he notes that better salaries alone won't protect against a potential scientific squeeze — Sykes says there has been a general drop in UK undergraduate enrolment in the sciences over the past few years. "There are other things that can attract people to places like Imperial," he says. Reputation, which you can't purchase, is one; infrastructure, which you can, is another.

But where do you find the cash to buy the tools you need to attract investigators in economically tough times? Equipment such as a 900-MHz nuclear magnetic resonance machine doesn't come cheap, even if its presence in a lab might help to lure in a world-class structural biologist. Sykes says that borrowing, begging and applying for grants are the only options.

Imperial has lately had success on some of those fronts. A £30-million donation by an American alumnus is helping to fund a new business school. And a recently announced £23.2-million loan from the European Investment Bank will help the college to upgrade its research and development facilities. But only time will tell if these investments will allow Imperial to attract more students and scientists to the campus.

Paul Smaglik
Naturejobs editor



Contents

POSTDOCS AND STUDENTS

Finding the right
postdoc position p784

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

ANNOUNCEMENTS

EVENTS



K. MEQUIER

Good track record with staff: Mildred Dresselhaus of the Massachusetts Institute of Technology with postdocs and students at a joint research group meeting.

Stacking the deck

Without doing the homework, choosing a postdoc position is a bit like picking a playing card at random. Karen Kreeger advises on how to come up trumps.

You're about to choose your first postdoc position, where you'll spend two or more years in a lab you may have never seen, with people you may have never met. How can you gain some knowledge about this important stop in your career?

Choosing the right lab and mentor can really affect your career, so it pays to do some preparation. The trick is to gather as much information as you can before you commit yourself to one place. Fortunately there are many ways to do this — using online resources, talking to past postdocs, interviewing present ones and, finally, visiting the lab.

Online research pays off: Erika Faselow and her mentor at Brown University, Barry Connors.

ON SCREENING

A good place to start is online. There aren't many formal, objective ways to judge a postdoc programme. But if you're looking in the United States, a survey by the US National Association of Graduate-Professional Students (NAGPS) due out this spring offers a chance to see what current graduate students and postdocs think of the programmes in which they are involved.

One advantage of the survey, says Audrey Elion, who is working on a PhD in psychology at Pennsylvania State University and is the legislative concerns committee coordinator for the NAGPS, is

that it guaranteed anonymity. As a result, she says, "students have been very open with direct comments about the programmes".

There are other, more indirect ways to judge a principal investigator online as well. When Erika Faselow, a neuroscience postdoc at Brown University in Providence, Rhode Island, was researching fellowships, she looked over abstracts for grants that had been made to potential principal investigators to ascertain the breadth of work and support for their labs. She found these abstracts using the Computer Retrieval of Information on Scientific Projects (CRISP) database, which contains federally funded biomedical research projects conducted at universities, hospitals and other research institutions and is managed by the US National Institutes of Health.

Once you've narrowed your choices down, the next step is to go to the prospective mentor's website and look in more detail at their research and publications, suggests Mildred Dresselhaus, institute professor of physics and electrical engineering at Massachusetts Institute of Technology in Cambridge. "I would recommend making a small investment in checking out the technical content," she advises. This way, you'll be able to see whether you will be learning about subject matter that interests you.

The next step is preparing to talk to principal





Speak to junior and senior lab members: Karen Spratt and her mentor Christopher Kemp at the Fred Hutchinson Cancer Research Center.

investigators and other people in their labs. Dresselhaus recommends a US National Academies' report from 2000, *Enhancing the Postdoctoral Experience for Scientists and Engineers*, as a good place to start; it includes a chapter on mentors.

Dresselhaus, who has had roughly 100 graduate students and 20 postdocs, sums up the report's advice: as well as checking the scientific-interest match, it is a good idea to look at the principal investigator's track record with staff. How long have people in the lab been there? How many postdocs and students have come and gone during the past couple of years?

Another question she suggests asking is, where have lab members landed jobs? You need to find out what became of people who were in the lab, not just one year later but also five years on. Ask if the principal investigator is open to any career path that you may choose — many academics try to steer all their protégés in the same direction. This may be especially important if you decide to go into industry or an alternative career such as writing or policy. Ask how the principal investigator has assisted postdocs in securing a job. Talk to more than one lab member: different personalities and relationships with the principal investigator will colour their comments.

Then there are questions about how research is conducted. Where is the lab's research headed? Will funding be available? Will you need to get your own?

Other types of questions relate to the atmosphere and working relationships in the lab. Does the principal investigator solicit postdocs' input, and listen to their ideas? Will the principal investigator discuss data with postdocs? To what extent do postdocs design experiments — in collaboration with the principal investigator or are they left to fend for themselves? Is it

a joint effort or does the principal investigator tell postdocs what to do? Do postdocs have a chance to work on their own projects? Will they have an opportunity to be first author on any papers?

Dresselhaus also suggests that prospective fellows ask themselves what they want to accomplish: for example, new skills, a new research direction, contacts, publications, conference attendances. Then ask principal investigators how they might help them achieve these goals.

TALK TO THEM

"The most important thing to do is talk to people," notes Deborah Stine, associate director for special projects at the US National Academy of Sciences. Conversations with potential mentors will help you make a better decision. But they're not the only people you should be talking to. Stine and others stress the need to unearth details about working conditions, as well as about the science done in the lab.

Where should you start asking about working conditions? Not at the top, says Karen Spratt, a postdoc at the Fred Hutchinson Cancer Research Center in Seattle. Instead, she recommends approaching junior lab members. "A lot of the time what you hear from the principal investigator is totally different from what you'll hear from the student or postdoc in the lab," she says.

It is a good idea to visit the lab before accepting a position, if possible, to get a feel for the way the principal investigator interacts with other members of the lab, as well as how the lab members work together.

It's not always what people say, but how and where, Spratt adds. "If you talk to someone in the lab and they say, 'Let's go down the hall,' that's a clue that there could be something wrong or some problem with other people in the lab."

Another way to learn about a prospective mentor, adds Fanselow, is to see how they interact with others at professional society meetings. "If they're giving a presentation, go to those," she says. "If they're giving a poster, watch them interacting with people to see what they're like."

This works both ways: when Fanselow gives a presentation, she takes note of how prospective colleagues relate to her. She also talks to past and present lab members at society meetings.

Even though publication record and scientific reputation are important factors, an equally important element is 'chemistry', or the way people interact, say Dresselhaus and Spratt.

Fanselow agrees about the importance of matching up the principal investigator's management style with the way you prefer to work. "Some people really like a hands-on approach and some people want to be left alone," she says. "It's probably hard to tell that, but you want to have a feel for whether they're going to let you do your own thing."

In the end, what makes the relationship between postdoc and principal investigator work is more about the subjective and intangible than the objective and measurable. But conducting background research, visiting labs and talking to past fellows can help prospective fellows take the temperature of both the subjective and the objective.

Karen Kreeger is a freelance science writer based in Philadelphia.

Web links

The 2000 National Doctoral Program Survey

♦ survey.nagps.org

Computer Retrieval of Information on Scientific Projects

♦ crisp.cit.nih.gov

Enhancing the Postdoctoral Experience for Scientists and Engineers

♦ www.nap.edu/catalog/9831.html